Temas de matemáticas

Cálculo diferencial de varias variables

Javier Páez Cárdenas

las prensas de ciencias

Javier Páez Cárdenas

Cálculo diferencial de varias variables

Facultad de Ciencias, UNAM
2017

Páez Cárdenas, Javier, autor. Cálculo diferencial de varias variables / Javier Páez Cárdenas.--1ª edición. -- Ciudad de México : Universidad Nacional Autónoma de México, Facultad de Ciencias, 2017.--(Temas de matemáticas)
XII, 506 páginas; 22 cm.

Incluye índice.
Bibliografía: página 499.
ISBN: 978-607-02-9887-5

1. Funciones de variables complejas. 2. Funciones de variables reales. 3. Cálculo diferencial. I. Universidad Nacional Autónoma de México. Facultad de Ciencias. II. Título. III. Serie.

515.9-scdd21 Biblioteca Nacional de México

Cálculo diferencial de varias variables
1a edición, 4 de noviembre de 2017
1a. reimpresión, enero de 2019

© DR. 2017. Universidad Nacional Autónoma de México
Ciudad Universitaria, 04510, Ciudad de México
Facultad de Ciencias
editoriales@ciencias.unam.mx

ISBN: 978-607-02-9887-5

Diseño de portada: *Laura Uribe y Eliete Martín del Campo*

Esta edición y sus características son propiedad de la Universidad Nacional Autónoma de México. Prohibida la reproducción total o parcial de la obra, por cualquier medio, sin la autorización por escrito del titular de los derechos.

Impreso y hecho en México

Contenido

Introducción vii

Agradecimientos xi

1. El conjunto $\mathbb{R}^n$ **1**
1.1. Para empezar, algunos ejemplos 1
1.2. Estructura algebraica de $\mathbb{R}^n$ 9
1.3. Aspectos geométricos de $\mathbb{R}^n$ 13
1.4. Otras normas 26
1.5. Topología de $\mathbb{R}^n$ 30
1.5.1. Clasificación de puntos 32
1.5.2. Conjuntos abiertos y cerrados 37
1.5.3. Otra clasificación de puntos 43
1.5.4. Conjuntos conexos 55
1.6. Otros sistemas coordenados 69
1.6.1. Coordenadas polares 70
1.6.2. Coordenadas cilíndricas 76
1.6.3. Coordenadas esféricas 80
1.7. Problemas 86

2. Funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ **95**
2.1. Álgebra y geometría de las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ 96
2.2. Límite y continuidad de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ 112
2.2.1. Sucesiones en $\mathbb{R}^n$ 112

2.2.2. Límite 123
2.2.3. Continuidad 135
2.2.4. Teoremas "fuertes" de continuidad 145
2.3. Continuidad uniforme 149
2.4. Problemas 161

3. La derivada de funciones de $\mathbb{R}$ en $\mathbb{R}^n$ 175
3.1. Geometría y movimiento 175
3.2. La derivada 182
3.3. Propiedades de la derivada 192
3.4. Derivada y geometría 198
3.5. Derivada y movimiento 218
3.6. Problemas 226

4. La derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}$ 237
4.1. Un interludio de Álgebra Lineal 237
4.2. La derivada direccional 248
4.2.1. Derivadas parciales 260
4.3. La derivada global 265
4.3.1. El gradiente 278
4.3.2. Otras propiedades de la derivada 286
4.3.3. La derivada en otras coordenadas 317
4.4. Derivadas direccionales de orden superior 328
4.5. Aproximación polinomial 338
4.6. Máximos y mínimos 351
4.6.1. Breve comentario sobre formas cuadráticas 374
4.6.2. Máximos y mínimos sobre restricciones 377
4.7. Problemas 397

5. La derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ 411
5.1. La derivada 411
5.1.1. Elementos básicos acerca de superficies 414
5.2. Propiedades de la derivada 421
5.2.1. Breve comentario sobre funciones coordenadas 427

5.3. La regla de la cadena 430
5.3.1. Cambio de coordenadas y regla de la cadena 433
5.4. El teorema de la función implícita 439
5.4.1. El caso lineal 439
5.4.2. El caso no lineal 445
5.5. El teorema de la función inversa 462
5.6. Problemas 488

Bibliografía **499**

Índice analítico **501**

Introducción

El tema principal del presente texto es el referente al concepto de derivada para funciones de varias variables, tema que constituye el núcleo central del curso de Cálculo Diferencial e Integral III que se imparte en la Facultad de Ciencias de la Universidad Nacional Autónoma de México.

En este trabajo partimos del supuesto de que el conjunto de conceptos y herramientas de Cálculo Diferencial e Integral que son del dominio del lector es aquel al que coloquialmente se le conoce como *Cálculo de una variable*. Como el lector seguramente también sabe, este nombre coloquial se debe a que los problemas y situaciones que dan origen a este conjunto de conocimientos son aquellos en los que *una* cantidad (o "variable") se puede poner en términos de *otra* cantidad, en donde cada una de estas "cantidades" tiene la particularidad de poderse "medir", "describir" o "representar" con un sólo número real. Esto mismo, dicho de una manera un poco más técnica, significa que el objeto matemático sobre el cual se construyen estos conceptos es el de *función de una variable real con valores reales*, o simplemente, *función de los reales en los reales*.

Una vez dicho lo anterior, es un poco más sencillo establecer cuál es el objetivo fundamental de este texto: desarrollar el concepto de derivada para el tipo de funciones en las que *una* cantidad (o variable) se puede poner en términos de *otra* cantidad, con la particularidad de que alguna de estas cantidades (¡o ambas!) no se puede medir, o representar con un sólo número real. Este tipo de funciones son conocidas en general como funciones de varias variables, y el conjunto de conceptos y resultados que desarrollaremos

alrededor de éstas es parte de lo que coloquialmente se conoce como *Cálculo de varias variables*.

Siguiendo este orden de ideas, los conceptos y resultados que se desarrollan en este texto se motivan principalmente a partir de los correspondientes conceptos y resultados para el caso de funciones de los reales en los reales. En virtud de esta característica, a lo largo de todo el texto se intenta usar un nivel adecuado de rigor y formalismo matemático. Por lo anterior, en este trabajo se parte del supuesto de que el lector conoce los temas básicos de Cálculo de una variable, de Geometría Analítica (plana y del espacio) y de Álgebra Superior. Por otra parte, dado que el primer curso de Cálculo de varias variables se suele tomar paralelamente con el primer curso de Álgebra Lineal, en este trabajo se intenta hacer uso de algunos conceptos de esta última materia una vez que el lector ya los haya estudiado.

Este trabajo está organizado en cinco capítulos; a continuación se da una somera y rápida descripción del contenido de cada uno de ellos.

De la misma forma que para realizar un estudio más profundo del Cálculo de una variable es necesario empezar por estudiar con mayor detenimiento al conjunto de los números reales, en el caso del Cálculo de varias variables es necesario hacer el mismo tipo de estudio del conjunto $\mathbb{R}^n$. Iniciamos el primer capítulo con una serie de ejemplos con los cuales se pretende mostrar que este conjunto resulta ser el más adecuado para representar a las variables ("independientes" y "dependientes") de las funciones para las cuales se definirán la mayoría de los conceptos de este texto. Inmediatamente después nos damos a la tarea de explorar las diferentes estructuras matemáticas que este conjunto posee: su estructura algebraica, su estructura geométrica y su estructura topológica. Concluimos este capítulo con un rápido repaso sobre los diferentes sistemas coordenados que se pueden establecer en el plano o en el espacio.

Una vez que en el capítulo 1 se estableció que las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ serán el principal objeto matemático con el cual podremos describir la mayoría de los problemas que dan lugar al estudio del Cálculo de varias variables, en el capítulo 2 comenzamos por hacer un rápido estudio de sus aspectos algebraicos y geométricos. Posteriormente nos damos a la tarea de introducir los conceptos de límite y continuidad de este tipo de funciones,

para lo cual previamente desarrollamos la herramienta de las sucesiones en $\mathbb{R}^n$. A continuación damos paso a los muy importantes teoremas "fuertes" de continuidad y concluimos este capítulo estudiando el concepto de continuidad uniforme.

El concepto de derivada para las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ es el tema central de este trabajo, y a este concepto dedicamos los restantes tres capítulos. En el capítulo 3 empezamos estudiando a las funciones de $\mathbb{R}$ en $\mathbb{R}^n$ y su particular importancia en la descripción de "objetos geométricos", y en la descripción del movimiento de un objeto en un plano o en el espacio. Basados en estas características, que entre otras ventajas permiten establecer cierta similitud con el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$, introducimos la derivada para este tipo de funciones y estudiamos su uso e interpretación justo en estos dos contextos: el geométrico y el cinemático.

La derivada para funciones de $\mathbb{R}^n$ en $\mathbb{R}$ es tal vez el primer concepto de este texto cuya definición resultará realmente novedosa para el lector. Esta característica se ve reflejada sobre todo en la necesidad de revisar previamente algunos conceptos importantes del Álgebra Lineal (bases ortonormales y funciones lineales) para poder dar dicha definición. Por esta razón, el capítulo 4 inicia con una revisión del concepto de base ortonormal, para después introducir la derivada direccional, que aún guarda una estrecha relación con la derivada de funciones de $\mathbb{R}$ en $\mathbb{R}$. Después de una breve revisión de la derivada de una función de $\mathbb{R}$ en $\mathbb{R}$ en términos de funciones lineales, se introduce el concepto de derivada (global) de una función de $\mathbb{R}^n$ en $\mathbb{R}$ y se prueban algunos resultados básicos.

Aun cuando la mayoría de los conceptos que se definen en el capítulo 4 se tratan de introducir de manera independiente de los sistemas coordenados más comunes, en este capítulo se incluye una sección en la que se deduce la forma específica que toman conceptos tales como el de gradiente, cuando las funciones con las que se está trabajando se expresan en sistemas coordenados diferentes a los euclideanos (en los casos de $\mathbb{R}^2$ y $\mathbb{R}^3$).

Como sucede en el caso de las funciones de $\mathbb{R}$ en $\mathbb{R}$, muchas de las aplicaciones prácticas del concepto de derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}$ están orientadas a la determinación de los valores máximos y mínimos de este tipo de funciones. Por esta razón (y algunas otras), se introducen los con-

ceptos de derivada direccional (y parcial) de orden superior y los polinomios de Taylor, los cuales son herramientas importantes para abordar estos problemas. Concluimos este capítulo con un análisis general del problema de máximos y mínimos, lo que nos conduce a un breve estudio de las formas cuadráticas (a fin de contar con una herramienta que nos permita clasificar los puntos críticos de una función), y a la formulación del Teorema de los multiplicadores de Lagrange, una de la herramientas más importantes que hay para la determinación de máximos y mínimos de funciones restringidas a ciertos conjuntos.

El capítulo 5, con el cual concluye este trabajo, aborda el tema de la derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$, cuya definición generaliza a las definiciones dadas en los capítulos 3 (para el caso $n = 1$) y 4 (para el caso $m = 1$). Una vez introducido este concepto, se prueban algunos resultados básicos relacionados con él, los cuales serán usados en la prueba de los teoremas más importantes de este capítulo: la Regla de la Cadena (en su versión más general), el Teorema de la Función Implícita (a partir del cual se prueba el Teorema de los multuplicadores de Lagrange) y el Teorema de la Función Inversa (a partir del cual se prueba el Teorema de la Función Implícita). Para la prueba de este último será necesario desarrollar una serie de resultados adicionales.

Al final de cada uno de estos capítulos se incluye una lista de problemas con los cuales se pretende que el lector refuerce, ponga a prueba y, en algunos casos, amplíe los conceptos estudiados.

Agradecimientos

Como suele suceder, detrás de la elaboración de una obra como la presente, no solo está el trabajo del autor, sino de muchas personas más. Sin duda una de las tareas más arduas y difíciles en el proceso de publicación de un texto tan amplio como el presente, es la lectura y revisión de sus primeras versiones. Por esta razón, agradezco profundamente a mi estimada colega Natalia Jonard, quien, estoy seguro, leyó y revisó las primeras versiones de este trabajo en más de una ocasión. Como es de suponerse, sus cuidadosas y rigurosas sugerencias y observaciones enriquecieron enormemente el contenido de este libro. Por la misma razón, también agradezco a otra querida colega, Emily Sánchez, y a un muy estimado exalumno, Carlos Armando Velázquez Fregoso, sus múltiples correcciones y observaciones las cuales también fueron muy importantes para el mejoramiento de este texto. Muchas personas más, colegas, alumnos y exalumnos, me hicieron muchas observaciones más. La lista es larga, y para no cometer penosas omisiones, les agradezco a todas ellas, de forma general y anónima. También agradezco a Ceci Neve su cuidadosa y refinada revisión de estilo, gracias a la cual este texto ahora se puede leer y entender mucho mejor. Por último, agradezco a Rafael Reyes el siempre tortuoso y largo trabajo de poner este texto en el formato de *Las Prensas de Ciencias*.

Capítulo 1

El conjunto $\mathbb{R}^n$

Así como en el caso del cálculo diferencial de una variable empezamos por estudiar al conjunto de los números reales $\mathbb{R}$, en este capítulo empezaremos por analizar más a fondo al conjunto $\mathbb{R}^n$, en virtud de que dicho conjunto será tanto el dominio como, en algunos casos, el contradominio de casi todas las funciones con las que trabajaremos en este texto. En la primera sección de este capítulo intentaremos mostrar porqué este tipo de conjuntos (para diferentes valores de n) es la forma más adecuada de representar al dominio y/o contradominio de la mayoría de las funciones que se encuentran en la base del cálculo de varias variables.

1.1. Para empezar, algunos ejemplos

Como mencionamos en la introducción, una motivación para el desarrollo del Cálculo de varias variables se puede encontrar en el planteamiento de ciertas situaciones en las que *una* cantidad (o variable, en general) se puede poner en términos de *otra* variable, con la particularidad de que alguna de estas variables (¡o ambas!) *no se puede medir, describir o representar con un sólo número real*. Lo que haremos en esta sección será presentar algunos ejemplos con estas características.

Ejemplo 1.1 *Supongamos que tenemos la suerte de contar con un dispositivo que nos permite saber, en un cierto instante, la temperatura en cada posición de la habitación en la que nos encontramos. Si queremos pensar esta situación en términos más técnicos, lo que estamos planteando es que este dispositivo nos permite establecer una función entre las distintas posiciones de la habitación y la temperatura en cada una de ellas.*

De esta forma, queda claro que una de nuestras variables (la independiente, como suele decirse) es la posición y la otra (la dependiente) es la temperatura.

Como seguramente el lector ya sabe, las diferentes posiciones en una habitación no son suceptibles de describirse con un sólo número real y con toda certeza también sabe que para representar cada posición nos harán falta tres números reales (e incluso también estará consciente de que para una misma posición dentro de la habitación, pueden haber diferentes ternas de números que la representan, ¡dependiendo del "sistema de referencia" que elijamos!).

Finalmente, la variable dependiente (en este caso la temperatura) sí se puede describir con un sólo número real, por lo que la situación que planteamos nos conduce a obtener una función que depende de tres números reales (los que describen o representan una posición) y que asigna otro número real (el que describe o mide la temperatura).

El ejemplo anterior permite una variante que nos lleva a obtener una función con una variable independiente diferente. Veamos de qué forma.

Ejemplo 1.2 *En el ejemplo anterior mencionamos que nuestro dispositivo nos permitía conocer la temperatura en cada posición de una habitación, en un cierto instante, situación que de inmediato nos hace pensar en una nueva variable: el instante en que estamos midiendo la temperatura.*

El lector estará de acuerdo en que esta nueva variable se puede medir con un sólo número real (y también estará de acuerdo en que a un mismo instante lo podrán describir diferentes números reales, ¡dependiendo en dónde colocamos el "instante cero"!).

De esta forma, considerando esa nueva variable, la función que obtenemos será una función cuya variable independiente estará dada por una

posición y un instante. Esta variable se puede describir con cuatro números reales (tres que sirven para describir o representar la posición, y uno para describir el instante). La variable dependiente estará dada por otro número real (el que describe o mide la temperatura).

Ejemplo 1.3 *Supongamos ahora que nos encontramos a la orilla de un río y que dentro de éste observamos una pequeña mota de polvo que se mueve como resultado de la corriente del mismo. Esta sencilla situación nos lleva de manera natural a pensar en una función: aquella que, para cada instante, nos da la posición de la mota de polvo.*

Como seguramente el lector ya dedujo, en este caso las dos variables importantes son: por un lado el tiempo (la variable independiente) que, como en el ejemplo anterior, podemos describir con un número real, y la posición (la variable dependiente), que como vimos en los ejemplos anteriores, se necesitan tres números reales (o dos, si el movimiento se realizara sobre un plano) para representarla.

Así pues, este ejemplo nos lleva a considerar una función que tiene como variable independiente al tiempo (que se puede medir con un sólo número real), y como "variable" dependiente a una posición, para la cual necesitamos tres cantidades (o números) para describirla.

Ejemplo 1.4 *Si todavía seguimos parados a la orilla del mismo río, podemos analizar la siguiente situación: si imaginamos al torrente del río como un conjunto de moléculas de agua que se están moviendo, pensemos en la función que, en un instante dado, nos asigna la velocidad con la que va viajando cada una de las moléculas que forman el torrente. De esta forma, nuestra función tendrá que ser tal que, a cada posición dentro del río, le asigna una velocidad.*

A estas alturas nos queda claro que a cada posición dentro del río (la variable independiente en este caso) la podemos describir con una terna de números reales, pero la velocidad (la variable dependiente) tal vez requiera un análisis aparte.

Cuando un objeto se mueve en línea recta es fácil convencerse de que su velocidad en un instante se puede medir por un sólo número real, pero este no

es el caso si el movimiento se da en un plano o en el espacio. Intuitivamente, si un objeto se mueve en un plano o en el espacio, su velocidad en un instante dado se puede representar por medio de una "flecha" cuyo punto inicial (u origen) se encuentra en la posición en la que se encuentra (en ese instante) el objeto en cuestión; la dirección de la "flecha" indicará la dirección del movimiento, y su "magnitud" (o longitud) indicará la rapidez con la que lo está haciendo.

Si ahora recordamos de nuestros cursos de Geometría Analítica que las flechas que parten de un punto fijo se pueden representar por una pareja (si estamos en el plano) o una terna (si estamos en el espacio) de números reales, tendremos que la variable dependiente de la función que describimos anteriormente (la velocidad de cada molécula, en un instante dado) se podrá describir por medio de una terna de números reales.

De esta forma, la función que obtenemos en este ejemplo asignará a una terna de números reales (la que describe la posición de una molécula en un cierto instante) otra terna de números reales (la que describe la velocidad de esa molécula en ese instante).

Como en el caso de ejemplo 1.1, el ejemplo anterior permite una variante que nos conduce a obtener una función con una variable independiente diferente.

Ejemplo 1.5 *Si en el ejemplo anterior, además de considerar la posición de cada molécula del agua del río en un instante dado, también consideramos diferentes instantes, entonces obtenemos una nueva función cuya variable independiente estará dada por una posición y un instante, la cual, como en el caso del ejemplo 1.2 se puede describir con cuatro números reales, y nos asignará una velocidad (su variable dependiente), que se podrá describir por tres números reales.*

En el último de nuestros ejemplos, mostraremos que las funciones (y sus variables) con las que nos podemos topar, pueden ser de muy diversa índole, y que éstas no siempre están relacionadas con posiciones, velocidades, tiempos o temperaturas.

Ejemplo 1.6 *Supongamos que en un laboratorio de investigación se encuentran realizando un experimento en el cual, para un valor de una cierta variable x, se obtiene un valor de otra variable y (vamos a suponer que los posibles valores de ambas variables se expresan con números reales).*

El experimento se realiza para k valores diferentes de la variable x, $x_1, \ldots, x_k$, ordenados de menor a mayor (es decir: $x_1 < \cdots < x_k$) y se obtienen k valores de la variable y, $y_1, \ldots, y_k$. De esta forma, se cuenta con k parejas de números reales $(x_1, y_1), \ldots, (x_k, y_k)$ las cuales se grafican como se muestra en la figura 1.1, (a). Si el gráfico que se obtiene (y el fenómeno con el que se está experimentando) sugieren que estos datos se deben parecer a (o "encajar" en) una recta, un problema importante es encontrar la recta que mejor se "ajuste" a estos datos.

En este momento no vamos a ver qué criterios son mejores para determinar si una recta se "ajusta" bien a un conjunto de datos, y sólo destacaremos que este problema nos conduce a considerar una función cuya variable independiente es una recta y cuya variable dependiente será (como veremos más adelante en el capítulo 4) un número real.

Aun y cuando a primera vista parezca un poco "exótico" eso de considerar una función cuya variable independiente sea una recta, no lo parecerá tanto si recordamos que toda recta (no vertical) en un plano tiene una ecuación de la forma $y = mx + b$, de tal manera que ésta queda totalmente determinada si conocemos m (su pendiente) y b (su ordenada al origen); es decir, toda recta (no vertical) en el plano se puede representar por medio de la pareja de números reales (m, b).

Tomando en consideración lo anterior, podemos concluir que la función a la que nos condujo este problema terminará siendo una que asociará a un par de números reales (m, b) (que en este caso "representan" a una recta, la variable independiente de la función) un número real.

Pero el ejemplo no termina aquí. ¿Cómo sería la función que tendríamos que considerar si lo que tenemos es que nuestro conjunto de datos se parece a (o encaja en) una parábola? (ver figura 1.1, (b)). Dado que las parábolas (con eje vertical) tienen en general una ecuación de la forma $y = a_0 + a_1 x + a_2 x^2$ (es decir, un polinomio de grado a lo más 2, que está determinado por sus coeficientes a_0, a_1 y a_2), la función a considerar sería entonces una que

asociaría a una terna de números reales (a_0, a_1, a_2) (la cual representa a una parábola si $a_2 \neq 0$, la variable independiente de nuestra función en este caso), un número real.

Como el lector seguramente ya está imaginando, podemos llevar este problema más lejos y considerar funciones que asocien a: "tétradas" (si nuestros datos se parecen a (o encajan en) la gráfica de un polinomio de grado a lo más 3), o a "quíntuplas" (si nuestros datos se parecen a la gráfica de un polinomio de grado a lo más 4), o en general, a "n-adas" (si nuestros datos se parecen a la gráfica de un polinomio de grado a lo más $n-1$), un número real.

Dicho de otra forma, nuestra función podría ser tal que su variable independiente fuera un polinomio de grado a lo más $n-1$ (los cuales se pueden representar por medio de "n-adas" de números reales).

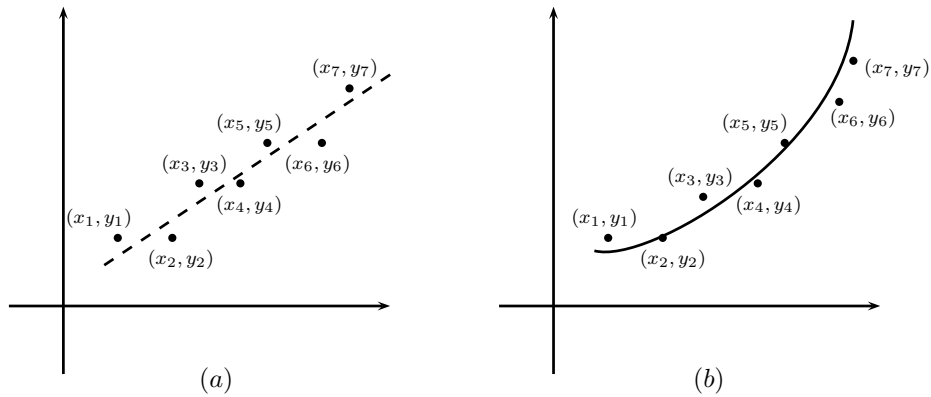


Figura 1.1: Los datos de un experimento, representados por cada pareja (x_i, y_i), sugieren que éstos "encajan" en una recta (a) o en una parábola (b).

Como resultado de esta larga lista de ejemplos, y de acuerdo con la intención original de presentarlos, resumiremos algunas de las características que tienen las variables (y/o "valores") de las funciones que describimos en ellos:

1. las variables (y/o valores) de estas funciones pueden ser de muy diversos tipos,

2. estas variables (y/o valores) siempre son susceptibles de representarse (describirse o medirse) por una cierta cantidad de números reales (dos, tres, cuatro ¡o más!),

3. esta representación no es única y en general depende del "*sistema de referencia*" que se elija.

Las características anteriores son muy importantes y algunas de ellas explican algunos de los términos que se suelen usar cuando nos referimos a la materia que nos ocupa, como es el caso del término: *Cálculo de varias variables*. Este término tiene su origen en la segunda característica que mencionamos, pues las diferentes cantidades que se necesitan para representar a la variable independiente de una función, también se consideran variables, de ahí que las funciones con las que se trabajará dependan de varias variables.

Otra importante observación que se debe hacer a partir de la segunda característica (sin duda la más importante), es que la representación de las variables (y/o valores) de estas funciones por medio de parejas, ternas, tétradas, o en general, "n-adas" de números reales, es un proceso que se suele realizar en el contexto de un concepto más amplio: el de *espacio vectorial*. El conjunto de posiciones" (o flechas que parten de un mismo punto) de un plano (o las correspondientes en el espacio), o el conjunto de polinomios de grado menor o igual a $n-1$ (que mencionamos en el ejemplo 1.6), son ejemplos particulares de este tipo de espacios.

Aunque el lector posiblemente todavía no esté muy familiarizado con este concepto, pronto aprenderá (en su curso de Álgebra Lineal I) que un espacio vectorial es un conjunto V que está dotado de dos operaciones: una suma y una multiplicación por escalares (cuando dichos escalares son números reales, decimos que V es un *espacio vectorial sobre los números reales*), operaciones a las cuales se le suelen pedir ciertas propiedades.

También pronto sabrá que, si $\beta = \{v_1, \ldots, v_n\}$ es un subconjunto de V que tiene la propiedad de que cualquier otro elemento $v \in V$ se puede

escribir de manera *única* como *combinación lineal* de los elementos de β (es decir, que existen $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, únicos, tales que $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$), entonces se dice que β es *una base* para V.

El concepto de base es muy importante, pues apoyándose en él es que se establece la representación de cada elemento $v \in V$ por medio de una "n-ada" de números reales $(\lambda_1, \ldots, \lambda_n)$; hacer esta representación por medio de diferentes bases es lo que está intimamente relacionado con los diferentes "*sistemas de referencia*" que mencionamos en el inciso 3. Cuando V tiene una base de este tipo, decimos que V es un espacio vectorial (sobre los números reales) de dimensión finita, específicamente de dimensión n.

De esta última observación, se desprende que las variables (independientes o dependientes) de las funciones con las que trabajaremos en este texto se pueden ver como elementos de un cierto espacio vectorial. Esto también explica por qué a todo este conjunto de conceptos y resultados relacionados con estas funciones, también se les conoce con el nombre de: *Cálculo vectorial*.

Por lo anterior, las funciones con las que vamos a trabajar deberían de estar consideradas, en general, como funciones definidas sobre un subconjunto de un espacio vectorial V y con contradominio sobre otro espacio vectorial W (ambos sobre los números reales y de dimensión finita).

Sin embargo, dado que cualquier espacio vectorial de este tipo se puede representar por medio del conjunto de "n-adas" de números reales (aunque esta representación no sea única y dependa de la *base* (o "*sistemas de referencia*") que se elija, lo que siempre habrá que recordar), a lo largo de este trabajo vamos a suponer que nuestras funciones estarán definidas sobre algún subconjunto de estas "n-adas", y tomarán sus valores (en general) sobre algún conjunto de "m-adas" (¡aunque se oiga un poco feo!).

A pesar de la observación anterior, en algunas ocasiones y para definir ciertos conceptos, no será necesario hacer referencia a la representación por medio de "n-adas" de los elementos de un cierto espacio vectorial. En estos casos usaremos letras del tipo $\hat{x}$ y $\hat{y}$ para denotar a estos elementos y escribiremos simplemente (no sin cometer cierto abuso de notación) $\hat{x} \in \mathbb{R}^n$ o $\hat{y} \in \mathbb{R}^n$.

Por lo dicho en los párrafos anteriores, lo siguiente que haremos será estudiar de manera más detallada al conjunto $\mathbb{R}^n$.

1.2. Estructura algebraica de $\mathbb{R}^n$

Pensar al conjunto $\mathbb{R}^n$ como una forma de representar a un espacio vectorial (sobre los reales y de dimensión n), tiene la ventaja de que la estructura algebraica de éste último se puede "exportar" o "trasladar" a $\mathbb{R}^n$. Esto es lo que nos proponemos hacer en esta sección. Antes de hacerlo, recordemos que el conjunto $\mathbb{R}^n$ está formado por las n-adas ordenadas $(x_1, \ldots, x_n)$ en donde cada x_i (a quien llamaremos la i-ésima coordenada de $(x_1, \ldots, x_n)$) es un número real, es decir

$$\mathbb{R}^n := \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{R}, i = 1, \ldots, n\}$$

y que también se suele decir que este conjunto es n veces el producto cruz (de conjuntos) de $\mathbb{R}$ consigo mismo.

A fin de motivar las operaciones que vamos a definir en $\mathbb{R}^n$, pensemos en el conjunto de todas las flechas del plano que comparten un punto inicial fijo, al que denotaremos por $\hat{0}$ y que llamaremos "*origen*". Es un hecho conocido que en este conjunto podemos definir (geométricamente y sin necesidad de recurrir a su representación por medio de parejas ordenadas) una operación de "suma" de "flechas", y una de "multiplicación" de un escalar real por una "flecha". Si estas operaciones tienen ciertas propiedades, permiten ver a dicho conjunto como un espacio vectorial.

Para definir la suma de las flechas (o vectores), utilizamos la llamada "ley del paralelogramo" que consiste en tomar el vector $\hat{y}$ y trasladar su punto inicial al punto final del vector $\hat{x}$, de tal forma que el vector que parte del origen y termina en el punto final del vector $\hat{y}$ (trasladado), será la flecha a la que llamaremos $\hat{x} + \hat{y}$ (ver figura 1.2, (a)).

Para la definición del producto de un escalar $\lambda \in \mathbb{R}$ por una flecha $\hat{x}$, que denotaremos por $\lambda\hat{x}$, tomamos cualquier "recta real" que pase por el origen $\hat{0}$ y que no "contenga" al vector $\hat{x}$, y realizamos la construcción de triángulos semejantes que se describe en la figura 1.2, (b))

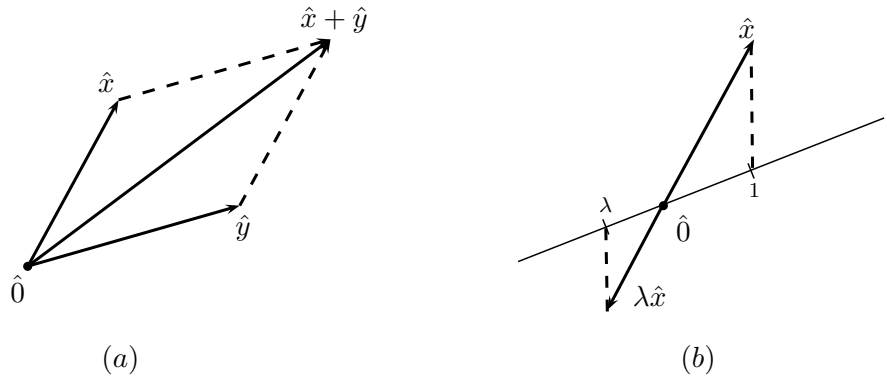


Figura 1.2: La suma (a) y el producto por un escalar (b) de vectores en $\mathbb{R}^2$.

Lo importante de haber hecho la definición geométrica de estas operaciones es que, si ahora establecemos un sistema de referencia que nos permita tener una representación de estas flechas en términos de parejas de números reales, obtenemos una forma de definir la multiplición por un escalar y suma de dichas parejas (seguramente el lector coincidirá en que es la forma más "natural" de definirlas).

Para obtener una base (o un sistema de referencia, que es como les vamos a llamar de aquí en adelante) que nos permita representar a cualquier flecha por medio de una pareja de números reales, eligiremos dos de estas flechas (que denotaremos por $\hat{v}_1$ y $\hat{v}_2$), en términos de las cuales haremos dicha representación. Es necesario que $\hat{v}_1$ y $\hat{v}_2$ no se encuentren sobre la misma recta, aunque el "ángulo" que formen no sea recto, ni sean de la misma longitud (ver figura 1.3).

Si ahora las parejas (x_1, x_2) y (y_1, y_2) representan, respectivamente, a la flecha $\hat{x}$ y a la flecha $\hat{y}$ en el sistema de referencia dado, lo que significa que

$$\hat{x} = x_1\hat{v}_1 + x_2\hat{v}_2 \qquad \text{y} \qquad \hat{y} = y_1\hat{v}_1 + y_2\hat{v}_2$$

y que simplemente expresaremos escribiendo que $\hat{x} = (x_1, x_2)$ y $\hat{y} = (y_1, y_2)$, uno puede comprobar (geométricamente) que las flechas $\hat{x} + \hat{y}$ y $\lambda\hat{x}$ estarán

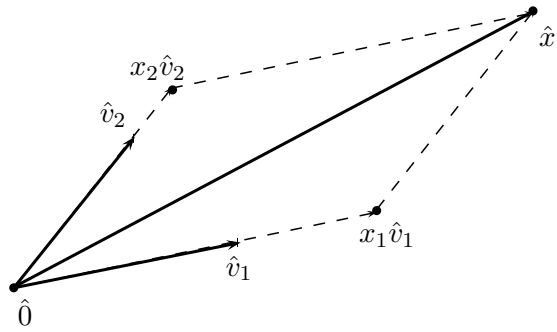


Figura 1.3: Significado geométrico del hecho de que la pareja (x_1, x_2) represente a la flecha (o vector) $\hat{x}$.

representadas, respectivamente, por las parejas (x_1+y_1, x_2+y_2) y $(\lambda x_1, \lambda x_2)$, es decir que

$$\hat{x} + \hat{y} = (x_1 + y_1, x_2 + y_2) \qquad \text{y} \qquad \lambda\hat{x} = (\lambda x_1, \lambda x_2).$$

Este procedimiento, que consiste en definir ciertos conceptos (en este caso la suma y producto por un escalar en el conjunto de las "flechas") partiendo de lo geométrico para después determinar cómo se expresan dichos conceptos en términos de la representación por n-adas (en este caso parejas), es un procedimiento al que recurriremos con mucha frecuencia para dotar al conjunto $\mathbb{R}^n$ de varias de sus "estructuras", tanto algebraicas como geométricas.

De hecho, con base en lo anterior definimos en el conjunto $\mathbb{R}^n$ un par de operaciones, una suma y un producto por escalares (reales), de la siguiente forma:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) := (x_1 + y_1, \ldots, x_n + y_n)$$

y

$$\lambda(x_1, \ldots, x_n) := (\lambda x_1, \ldots, \lambda x_n)$$

para $\lambda \in \mathbb{R}$.

Es muy fácil probar que todas las propiedades que tiene la suma de números reales, se "heredan" a esta suma definida en $\mathbb{R}^n$. En particular,

a la n-ada cuyas coordenadas son todas cero lo llamaremos *el origen* y la denotaremos por $\hat{0}$, es decir

$$\hat{0} := (0, \ldots, 0),$$

del mismo modo que si $\hat{x} = (x_1, \ldots, x_n)$, denotaremos por $-\hat{x}$ a la n-ada $(-x_1, \ldots, -x_n)$, es decir

$$-\hat{x} := (-x_1, \ldots, -x_n) = (-1)(x_1, \ldots, x_n).$$

A su vez, aprovecharemos para definir la resta de elementos en $\mathbb{R}^n$ (la cual, como en el caso de $\mathbb{R}$, no es mas que una suma "encubierta"), de la siguiente manera: si $\hat{x}, \hat{y} \in \mathbb{R}^n$ definimos

$$\hat{x} - \hat{y} := \hat{x} + (-\hat{y}).$$

Si pensamos a $\hat{x}$ y $\hat{y}$ como flechas en el plano o en el espacio, la flecha $\hat{x} - \hat{y}$ coincide con ser la flecha que se obtiene uniendo el punto final de $\hat{y}$ con el punto final de $\hat{x}$, trasladada al origen, como se muestra en la figura 1.4 para el caso del plano.

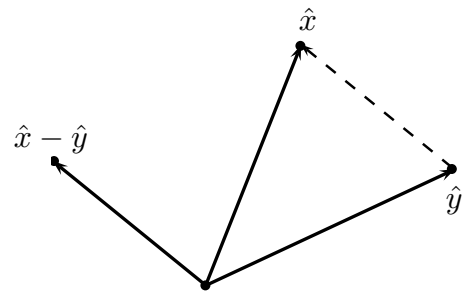


Figura 1.4: Construcción geométrica del vector $\hat{x} - \hat{y}$

Para concluir esta breve sección, simplemente resaltaremos el hecho de que en $\mathbb{R}^n$ existen dos propiedades distributivas: el producto por un escalar distribuye a la suma de elementos en $\mathbb{R}^n$, y el producto de un elemento de $\mathbb{R}^n$ distribuye a la suma de escalares, es decir, si $\hat{x}, \hat{y} \in \mathbb{R}^n$ y $\lambda, \gamma \in \mathbb{R}$, entonces

$$\lambda\,(\hat{x} + \hat{y}) = \lambda\hat{x} + \lambda\hat{y} \quad \text{y} \quad (\lambda + \gamma)\,\hat{x} = \lambda\hat{x} + \gamma\hat{x}.$$

1.3. Aspectos geométricos de $\mathbb{R}^n$

Como vimos en la sección anterior, para establecer una correspondencia entre los puntos o flechas del plano con las parejas ordenadas de números reales, no es necesario elegir un sistema de referencia en el que las flechas que lo formen tengan que ser perpendiculares ni de la misma longitud. Tomarlas de esta forma, en cuyo caso diremos que nuestro sistema de referencia es un *sistema coordenado cartesiano*[1], es una libertad adicional que nos podemos dar, y que nos permitirá "trasladar" al conjunto $\mathbb{R}^n$ toda la estructura geométrica que este concepto de perpendicularidad conlleva.

De esta forma, cuando usamos a la pareja (x_1, x_2) o a la terna (x_1, x_2, x_3) de números reales para designar a una flecha (o vector) de un cierto sistema coordenado cartesiano, la longitud (o magnitud) de esta flecha es una cantidad que podemos escribir en términos de las coordenadas correspondientes. La expresión algebraica que representa a esta cantidad se deduce fácilmente usando el Teorema de Pitágoras (como se muestra en la figura 1.5), y está dada por

$$\sqrt{x_1^2 + x_2^2}$$

para el caso del plano ($\mathbb{R}^2$), y por

$$\sqrt{x_1^2 + x_2^2 + x_3^2}$$

para el caso del espacio ($\mathbb{R}^3$). Es importante resaltar que si la pareja o la terna designa a un punto (en lugar de una flecha), lo que estas expresiones representan es la distancia que hay entre dicho punto y el origen.

Tomando como base las expresiones anteriores, a cada $\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ (para cualquier $n \in \mathbb{N}$) le asociaremos un número real positivo, al que llamaremos *la norma (euclideana) de* $\hat{x}$, y que denotaremos por $\|\hat{x}\|$. Este nuevo concepto está definido de la siguiente manera.

[1]Nombrado así en recuerdo de René Descartes (La Haye, Turena francesa, 31 de marzo de 1596 - Estocolmo, Suecia, 11 de febrero de 1650), también llamado Renatus Cartesius, quien fue un filósofo, matemático y físico francés, considerado como el padre de la geometría analítica y de la filosofía moderna, así como uno de los nombres más destacados de la revolución científica. (fuente: Wikipedia).

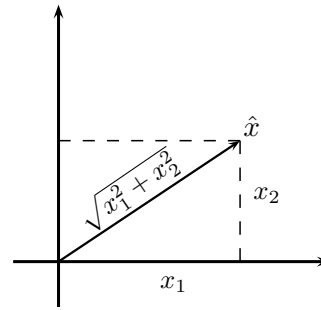


Figura 1.5: La cantidad $\sqrt{x_1^2 + x_2^2}$ representa la magnitud del vector $\hat{x} = (x_1, x_2)$.

Definición 1.1 *Para cada $\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ definimos la norma (euclideana) de $\hat{x}$, que denotamos por $\|\hat{x}\|$, como*

$$\|\hat{x}\| := \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{1.1}$$

Aun cuando n sea mayor que 3, de acuerdo con la interpretación como una distancia que se le da a dicha expresión para el caso en que $n = 2$ o $n = 3$, en general diremos que el número $\|\hat{x}\|$ representa *la distancia* entre el "punto" determinado por $\hat{x}$ y el origen $\hat{0}$.

A fin de explorar las propiedades del concepto que acabamos de definir, vale la pena hacer notar que si en esta definición tomamos $n = 1$ (en cuyo caso nuestro conjunto coincide con ser $\mathbb{R}$), entonces la norma no es otra cosa más que el conocidísimo concepto de valor absoluto de los números reales, lo que por cierto, nos permite pensar al concepto de norma como una generalización a $\mathbb{R}^n$ del correspondiente concepto de valor absoluto de los números reales. Es precisamente a partir de este hecho, y recordando las propiedades más elementales del valor absoluto, que podemos establecer la siguiente

Proposición 1.1 *La norma (euclideana) satisface las siguientes propiedades:*

1. $\|\hat{x}\| \geq 0$ para toda $\hat{x} \in \mathbb{R}^n$ y $\|\hat{x}\| = 0$ si y sólo si $\hat{x} = \hat{0}$
2. $\|\lambda\hat{x}\| = |\lambda|\,\|\hat{x}\|$ para toda $\hat{x} \in \mathbb{R}^n$ y para toda $\lambda \in \mathbb{R}$
3. $\|\hat{x} + \hat{y}\| \leq \|\hat{x}\| + \|\hat{y}\|$ para cualesquiera $\hat{x}, \hat{y} \in \mathbb{R}^n$ (desigualdad del triángulo)

Demostración. Las afirmaciones de los incisos 1 y 2 son inmediatas. La prueba del inciso 3 requiere un nuevo concepto que desarrollaremos a continuación. ■

Para poder probar el inciso 3 de la proposición 1.1 introduciremos un nuevo concepto que vamos a motivar a partir de un problema geométrico muy sencillo: dados dos vectores $\hat{x}, \hat{y} \in \mathbb{R}^2$, distintos de $\hat{0}$, ¿cuál es el escalar $\lambda \in \mathbb{R}$ que hace que los vectores $\hat{x} - \lambda\hat{y}$ y $\hat{y}$ sean perpendiculares? (ver figura 1.6).

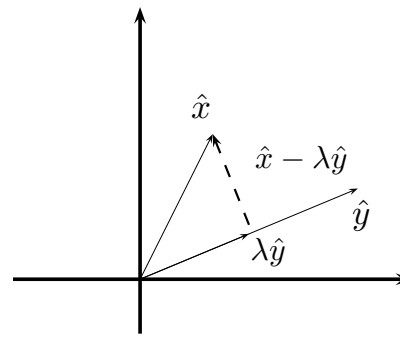


Figura 1.6: ¿Cuál es el valor de λ que hace que los vectores $\hat{x} - \lambda\hat{y}$ y $\hat{y}$ sean perpendiculares?

Supongamos que los vectores $\hat{x}$ y $\hat{y}$ tienen coordenadas (x_1, x_2) y (y_1, y_2), respectivamente. Dado que deseamos que el triángulo formado por los vectores $\hat{x}, \hat{x} - \lambda\hat{y}$ y $\lambda\hat{y}$ sea un triángulo rectángulo, por el teorema de Pitágoras debemos tener que

$$\|\lambda\hat{y}\|^2 + \|\hat{x} - \lambda\hat{y}\|^2 = \|\hat{x}\|^2 \tag{1.2}$$

lo cual, escrito en términos de las coordenadas de los vectores, se traduce en

$$\left(\sqrt{(\lambda y_1)^2 + (\lambda y_2)^2}\right)^2 + \left(\sqrt{(x_1 - \lambda y_1)^2 + (x_2 - \lambda y_2)^2}\right)^2 = \left(\sqrt{x_1^2 + x_2^2}\right)^2.$$

Cancelando los cuadrados con las raíces cuadradas, desarrollando los cuadrados que se encuentran dentro de la segunda raíz cuadrada y cancelando y factorizando los términos iguales, llegamos a que λ debe satisfacer la ecuación

$$2\lambda^2 \left(y_1^2 + y_2^2\right) - 2\lambda \left(x_1 y_1 + x_2 y_2\right) = 0,$$

la cual tiene las soluciones $\lambda = 0$ y

$$\lambda = \frac{x_1 y_1 + x_2 y_2}{y_1^2 + y_2^2} = \frac{x_1 y_1 + x_2 y_2}{\|\hat{y}\|^2}. \tag{1.3}$$

Es importante observar que la solución $\lambda = 0$ no tiene mucho que ver con el problema planteado, puesto que, independientemente de la posición de $\hat{x}$ y $\hat{y}$, tomando este valor de λ siempre se satisface la ecuación 1.2. Por esta razón, la segunda solución de la ecuación es la importante para el problema que planteamos.

La expresión $x_1 y_1 + x_2 y_2$ (que aparece en el numerador de la expresión del número λ que calculamos renglones arriba) resultará ser muy relevante a la hora de precisar un concepto que hasta ahora sólo podemos manejar de una forma muy intuitiva: el ángulo formado por los vectores $\hat{x}$ y $\hat{y}$.

Podemos reforzar la sospecha de que existe una relación entre esta expresión y el concepto de ángulo si en el problema que planteamos anteriormente observamos que, si los vectores $\hat{x}$ y $\hat{y}$ ya fueran perpendiculares, la única solución sería $\lambda = 0$ de donde se concluye que la expresión $x_1 y_1 + x_2 y_2$ tendría que ser 0. Esto se puede confirmar con algunos ejemplos, como sería el caso de los vectores $(1, 0)$ y $(0, 1)$, que de acuerdo con el sistema coordenado con el que estamos trabajando, se tomaron de tal forma que fueran perpendiculares. Seguramente el lector podrá verificar en muchos más casos específicos que la perpendicularidad de dos vectores se corresponde con el hecho de que la expresión $x_1 y_1 + x_2 y_2$ vale 0).

También es fácil verificar que, dado el vector $\hat{x} = (x_1, x_2) \neq \hat{0}$ se tiene que el vector $\hat{y} = (-x_2, x_1) \neq \hat{0}$ es perpendicular a $\hat{x}$ puesto que junto con el vector $\hat{x} - \hat{y}$ forman un triángulo rectángulo, lo que se deduce del hecho de que dicho triángulo satisface el teorema de Pitágoras:

$$\begin{aligned} \|\hat{x} - \hat{y}\|^2 &= \|(x_1, x_2) - (-x_2, x_1)\|^2 \\ &= (x_1 + x_2)^2 + (x_2 - x_1)^2 \\ &= \left(x_1^2 + x_2^2\right) + \left((-x_2)^2 + x_1^2\right) \\ &= \|\hat{x}\|^2 + \|\hat{y}\|^2 \end{aligned}$$

(ver figura 1.7).

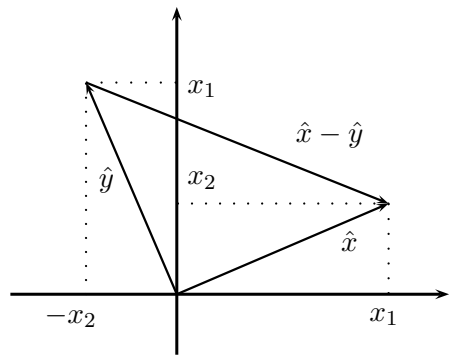


Figura 1.7: Los vectores $\hat{x} = (x_1, x_2)$ y $\hat{y} = (-x_2, x_1)$ son perpendiculares.

Por otra parte, es fácil ver que en este caso también se tiene que

$$\begin{aligned} x_1y_1 + x_2y_2 &= x_1(-x_2) + x_2x_1 \\ &= 0. \end{aligned}$$

Además de sospechar que la perpendicularidad entre los vectores $\hat{x}$ y $\hat{y}$ está íntimamente relacionada con el hecho de que la expresión $x_1y_1 + x_2y_2$ es igual a cero, de la identidad 1.3 y las figuras 1.8 deducimos que esta

expresión debe ser positiva si el ángulo formado por los vectores es agudo (mayor que 0 y menor que $\pi/2$), y negativa si el ángulo formado por los vectores es obtuso (mayor que $\pi/2$ y menor que π).

Con el fin terminar de reforzar nuestras sospechas sobre la relación que esta expresión tiene con el concepto de "ángulo" entre dos vectores, ahora vamos a resolver el problema planteado inicialmente recurriendo justamente a este concepto, y a las funciones trigonométricas básicas que se definen con base en éste.

Si θ es el ángulo formado por los vectores $\hat{x}$ y $\hat{y}$, y éste es agudo (es decir que $0 < \theta < \pi/2$), sabemos que $\lambda > 0$ (ver figura 1.8 (a)) y que

$$\begin{aligned}\cos(\theta) &= \frac{\|\lambda\hat{y}\|}{\|\hat{x}\|} \\ &= \frac{\lambda\,\|\hat{y}\|}{\|\hat{x}\|}.\end{aligned}$$

Si θ es un ángulo obtuso (es decir que $\pi/2 < \theta < \pi$), sabemos que $\lambda < 0$ (ver figura 1.8 (b)) y que

$$\begin{aligned}-\cos(\theta) &= \cos(\pi - \theta) \\ &= \frac{\|\lambda\hat{y}\|}{\|\hat{x}\|} \\ &= \frac{-\lambda\,\|\hat{y}\|}{\|\hat{x}\|},\end{aligned}$$

de tal forma que en ambos casos obtenemos que

$$\lambda = \frac{\|\hat{x}\|\cos(\theta)}{\|\hat{y}\|}.$$

Si ahora igualamos los valores que hemos obtenido de λ por estos dos caminos, tenemos que

$$\frac{x_1y_1 + x_2y_2}{\|\hat{y}\|^2} = \frac{\|\hat{x}\|\cos(\theta)}{\|\hat{y}\|},$$

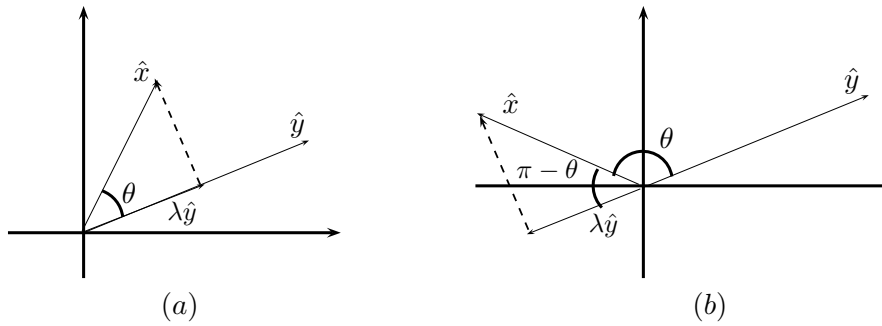


Figura 1.8: El cálculo de λ usando el ángulo θ.

de donde

$$\cos(\theta) = \frac{x_1y_1 + x_2y_2}{\|\hat{x}\| \, \|\hat{y}\|}. \tag{1.4}$$

Esta última identidad es sin duda toda una revelación en virtud de que nos proporciona una forma más específica y rigurosa de medir al ángulo formado por los vectores $\hat{x}$ y $\hat{y}$ en términos de sus coordenadas. En efecto, de la igualdad anterior obtenemos que

$$\begin{aligned} \theta &= \cos^{-1}\left(\frac{x_1y_1 + x_2y_2}{\|\hat{x}\| \, \|\hat{y}\|}\right) \\ &= \arccos\left(\frac{x_1y_1 + x_2y_2}{\|\hat{x}\| \, \|\hat{y}\|}\right), \end{aligned} \tag{1.5}$$

en donde elegimos la "rama" de la función arc cos que toma sus valores en el intervalo $[0, \pi]$.

Que la identidad anterior es una buena forma de medir dicho ángulo se confirma si observamos, por ejemplo, que este número no varía, como es de esperarse, si tomamos otro par de vectores que apunten en las mismas direcciones en las que apuntan los vectores $\hat{x}$ y $\hat{y}$, respectivamente. Esta condición se traduce en tomar vectores de la forma $\lambda\hat{x}$ y $\mu\hat{y}$, con $\lambda, \mu > 0$; en este caso se tiene que, si $\hat{x} = (x_1, x_2)$ y $\hat{y} = (y_1, y_2)$, entonces $\lambda\hat{x} = (\lambda x_1, \lambda x_2)$

y $\mu\hat{y} = (\mu y_1, \mu y_2)$ de tal forma que si llamamos $\tilde{\theta}$ al ángulo formado por los vectores $\lambda\hat{x}$ y $\mu\hat{y}$ y lo calculamos en términos de sus coordenadas, de acuerdo con la ecuación 1.5, se tiene que

$$\begin{aligned}\tilde{\theta} &= \cos^{-1}\left(\frac{(\lambda x_1)(\mu y_1) + (\lambda x_2)(\mu y_2)}{\|\lambda\hat{x}\|\,\|\mu\hat{y}\|}\right)\\ &= \cos^{-1}\left(\frac{\lambda\mu(x_1y_1 + x_2y_2)}{\lambda\mu\,\|\hat{x}\|\,\|\hat{y}\|}\right)\\ &= \cos^{-1}\left(\frac{x_1y_1 + x_2y_2}{\|\hat{x}\|\,\|\hat{y}\|}\right)\\ &= \theta.\end{aligned}$$

Nótese que este último hecho nos permite suponer que los vectores $\hat{x}$ y $\hat{y}$ tienen norma 1, puesto que si este no fuera el caso, bastaría con tomar $\lambda = 1/\|\hat{x}\|$ y $\mu = 1/\|\hat{y}\|$ para que los vectores $\lambda\hat{x}$ y $\mu\hat{y}$ sí tuvieran norma 1. Por cierto que, cuando a un vector $\hat{x} \neq \hat{0}$ le aplicamos el procedimiento de multiplicarlo por el recíproco de su norma, diremos que hemos *"normalizado"* al vector $\hat{x}$; es decir, *"normalizar"* al vector $\hat{x}$ significará tomar el vector $(1/\|\hat{x}\|)\hat{x}$, el vector que está en la misma dirección que $\hat{x}$ y cuya norma es 1). Recurriremos con tanta frecuencia a este procedimiento de normalización, que en lugar de escribir $(1/\|\hat{x}\|)\hat{x}$ simplemente escribiremos $\hat{x}/\|\hat{x}\|$.

Resumiendo toda la discusión anterior y tomando en cuenta las identidades 1.4 y 1.5, podemos concluir que la expresión $x_1y_1 + x_2y_2$ contiene la información suficiente como para poder calcular el ángulo formado por los vectores $\hat{x}$ y $\hat{y}$, en donde (x_1, x_2) es la pareja que representa (en un sistema coordenado cartesiano) a $\hat{x}$ y (y_1, y_2) la que representa a $\hat{y}$. Por esta razón, usaremos dicha expresión para definir sobre el conjunto $\mathbb{R}^n$ otra operación (a la que llamaremos *"el producto punto"* (o *"el producto interior"*)) entre dos elementos $\hat{x}, \hat{y} \in \mathbb{R}^n$, y que denotaremos por $\hat{x} \cdot \hat{y}$.

Definición 1.2 *Dados $\hat{x}, \hat{y} \in \mathbb{R}^n$, con $\hat{x} = (x_1, \ldots, x_n)$ y $\hat{y} = (y_1, \ldots, y_n)$, definimos el producto punto (o producto interior) de $\hat{x}$ y $\hat{y}$, que denotaremos*

por $\hat{x} \cdot \hat{y}$*, como el número real dado por* $x_1y_1 + \cdots + x_ny_n$*, es decir*

$$\begin{aligned}\hat{x} \cdot \hat{y} &:= x_1y_1 + \cdots + x_ny_n \\ &:= \sum_{i=1}^{n} x_iy_i.\end{aligned}$$

Esta nueva operación tiene una serie de propiedades básicas, las cuales resumimos en la siguiente

Proposición 1.2 *Dados* $\hat{x}, \hat{y}, \hat{z} \in \mathbb{R}^n$ *y* $\lambda \in \mathbb{R}$*, el producto punto de vectores definido en 1.2 satisface las siguientes propiedades:*

1. $\hat{x} \cdot \hat{x} \geq 0$ *y* $\hat{x} \cdot \hat{x} = 0$ *si y sólo si* $\hat{x} = \hat{0}$
2. $\hat{x} \cdot \hat{y} = \hat{y} \cdot \hat{x}$
3. $\hat{x} \cdot (\hat{y} + \hat{z}) = \hat{x} \cdot \hat{y} + \hat{x} \cdot \hat{z}$
4. $(\lambda\hat{x}) \cdot \hat{y} = \hat{x} \cdot (\lambda\hat{y}) = \lambda(\hat{x} \cdot \hat{y})$

Además de mencionar que la demostración de estas propiedades es muy sencilla (y que por esta razón se deja como ejercicio al lector), es importante señalar que a cualquier otra función definida de $\mathbb{R}^n \times \mathbb{R}^n$ en $\mathbb{R}$ que las satisfaga, también se le llamará un producto punto (o producto interior).

Si bien es cierto que es importante mostrar que la función dada en la definición 1.2 satisface las propiedades de la proposición anterior (pues con ello se prueba que dicha función sí es un producto punto), también es importante destacar la íntima relación que existe entre este producto y el concepto de norma euclideana que definimos anteriormente, la cual queda expresada en la identidad

$$\|\hat{x}\|^2 = \hat{x} \cdot \hat{x}. \tag{1.6}$$

Pero sin lugar a dudas, la propiedad más importante que hay que destacar de esta nueva operación, está relacionada con la identidad 1.4. En efecto,

si escribimos esa misma identidad usando la notación del producto punto, esta se traduce en la siguiente identidad:

$$\cos(\theta) = \frac{\hat{x} \cdot \hat{y}}{\|\hat{x}\| \, \|\hat{y}\|} \tag{1.7}$$

la cual, por ahora, sólo podemos dar por cierta para el caso de $\mathbb{R}^2$, y sin ignorar que su demostración se obtiene a partir de nuestro (no muy bien definido) concepto de ángulo y su relación con las funciones trigonométricas (las que por cierto, si están bien definidas).

La conclusión más importante que podemos obtener de la identidad 1.7 se deduce del hecho de que los valores de la función coseno se encuentran entre -1 y 1. En efecto, tomando en cuenta esto, tendremos que para cualesquiera par de vectores $\hat{x}, \hat{y} \in \mathbb{R}^2$ diferentes de $\hat{0}$ se debe cumplir que

$$\frac{|\hat{x} \cdot \hat{y}|}{\|\hat{x}\| \, \|\hat{y}\|} \leq 1$$

o que

$$|\hat{x} \cdot \hat{y}| \leq \|\hat{x}\| \, \|\hat{y}\| \, . \tag{1.8}$$

Nótese que si alguno de estos vectores fuera $\hat{0}$, entonces se debería satisfacer la igualdad.

Lo interesante de la desigualdad 1.8 es que no involucra a ningún ángulo y tiene todo el sentido preguntarse si es válida en cualquier $\mathbb{R}^n$. ¡Y lo mejor de todo es que la respuesta a esta pregunta es positiva! En efecto, la desigualdad 1.8 es válida para cualquier par de vectores en $\mathbb{R}^n$ y es conocida como la *desigualdad de Cauchy-Schwarz*[2]. Esta desigualdad jugará un papel muy importante a lo largo de todo este texto, y por esta misma razón le concederemos el nivel de teorema.

[2]Llamada así en honor del matemático francés Augustin Louis Cauchy (París, 21 de agosto de 1789 - Sceaux, 23 de mayo de 1857), quien la publicó en 1821, y del matemático alemán Karl Hermann Amandus Schwarz (25 de enero 1843 - 30 de noviembre 1921), quien la publicó en 1888. Aunque dicha desigualdad también fue establecida por el matemático ruso Viktor Yakovlevich Bunyakovsky (Bar, Ucrania, 16 de diciembre 1804 - San Petesburgo, Rusia, 12 de diciembre 1889) en 1859. (fuente: Wikipedia).

Teorema 1.1 (Desigualdad de Cauchy-Schwarz) *Para cualesquiera par de vectores $\hat{x}, \hat{y} \in \mathbb{R}^n$ se satisface que*

$$|\hat{x} \cdot \hat{y}| \leq \|\hat{x}\| \, \|\hat{y}\| .$$

Demostración. Como ya se mencionó, esta desigualdad es inmediata si $\hat{x}$ o $\hat{y}$ es el vector $\hat{0}$. Por esta razón, supondremos que $\hat{x}$ y $\hat{y}$ son distintos de $\hat{0}$. Asimismo, probar la desigualdad del enunciado es equivalente a probar que

$$\left| \left(\frac{\hat{x}}{\|\hat{x}\|} \right) \cdot \left(\frac{\hat{y}}{\|\hat{y}\|} \right) \right| = \frac{|\hat{x} \cdot \hat{y}|}{\|\hat{x}\| \, \|\hat{y}\|} \leq 1,$$

por lo que también podemos suponer que

$$\|\hat{x}\| = \|\hat{y}\| = 1.$$

Una vez establecido lo anterior, por los incisos 1 y 3 de la proposición 1.2 y la identidad 1.6, sabemos que

$$\begin{aligned} 0 &\leq (\hat{x} - \hat{y}) \cdot (\hat{x} - \hat{y}) \\ &= \hat{x} \cdot \hat{x} - 2\hat{x} \cdot \hat{y} + \hat{y} \cdot \hat{y} \\ &= 2(1 - \hat{x} \cdot \hat{y}), \end{aligned}$$

de donde concluimos que

$$\hat{x} \cdot \hat{y} \leq 1.$$

Con base en las mismas propiedades, también sabemos que

$$\begin{aligned} 0 &\leq (\hat{x} + \hat{y}) \cdot (\hat{x} + \hat{y}) \\ &= \hat{x} \cdot \hat{x} + 2\hat{x} \cdot \hat{y} + \hat{y} \cdot \hat{y} \\ &= 2(1 + \hat{x} \cdot \hat{y}) \end{aligned}$$

y ahora concluimos que

$$-1 \leq \hat{x} \cdot \hat{y},$$

con lo que hemos probado que

$$-1 \leq \hat{x} \cdot \hat{y} \leq 1,$$

o equivalentemente, que

$$|\hat{x} \cdot \hat{y}| \leq 1,$$

que es lo que deseábamos demostrar. ■

Es importante llamar la atención del lector sobre el hecho de que en esta prueba no hubo necesidad de recurrir a las coordenadas de los vectores $\hat{x}$ y $\hat{y}$, y que las herramientas importantes fueron las propiedades básicas del producto punto y su relación con la norma. Lo relevante de esta observación es que, si definimos otro producto punto sobre el conjunto $\mathbb{R}^n$ que satisfaga las propiedades de la proposición 1.2, y con base en este otro producto punto definimos otra norma por medio de la identidad 1.6, esta otra norma cumplirá las propiedades de la proposición 1.1 y la desigualdad de Cauchy-Schwarz, usando ese otro producto punto y esa otra norma, ¡seguirá siendo cierta!

Como una primera muestra de la importancia y utilidad de la desigualdad de Cauchy-Schwarz, ahora la usaremos para probar el inciso 3 de la proposición 1.1 (la desigualdad del triángulo), junto con otra desigualdad que también nos resultará muy útil a lo largo de este texto.

Proposición 1.3 *Para cualesquiera par de vectores $\hat{x}, \hat{y} \in \mathbb{R}^n$ se satisface que:*

1. $\|\hat{x} + \hat{y}\| \leq \|\hat{x}\| + \|\hat{y}\|$ *(desigualdad del triángulo)*
2. $|\|\hat{x}\| - \|\hat{y}\|| \leq \|\hat{x} + \hat{y}\|$.

Demostración. Para la prueba de la desigualdad del triángulo (inciso 1), con base en la identidad 1.6 y la desigualdad de Cauchy-Schwarz, tenemos que

$$\begin{aligned}
\|\hat{x} + \hat{y}\|^2 &= (\hat{x} + \hat{y}) \cdot (\hat{x} + \hat{y}) \\
&= \hat{x} \cdot \hat{x} + 2\hat{x} \cdot \hat{y} + \hat{y} \cdot \hat{y} \\
&= \|\hat{x}\|^2 + 2\hat{x} \cdot \hat{y} + \|\hat{y}\|^2 \\
&\leq \|\hat{x}\|^2 + 2\,\|\hat{x}\|\,\|\hat{y}\| + \|\hat{y}\|^2 \qquad \text{(desigualdad de Cauchy-Schwarz)} \\
&= (\|\hat{x}\| + \|\hat{y}\|)^2,
\end{aligned}$$

de modo que, sacando raíz cuadrada, llegamos a la desigualdad deseada.

Para la prueba del inciso 2, dado que

$$\|\hat{x}\| = \|(\hat{x} + \hat{y}) + (-\hat{y})\|,$$

usando el inciso anterior (la desigualdad del triángulo) y el hecho de que $\|-\hat{y}\| = \|(-1)\hat{y}\| = \|\hat{y}\|$ (inciso 2 de la proposición 1.1), se tiene que

$$\begin{aligned} \|\hat{x}\| &= \|(\hat{x} + \hat{y}) + (-\hat{y})\| \\ &\leq \|\hat{x} + \hat{y}\| + \|\hat{y}\| \end{aligned}$$

y por lo tanto

$$\|\hat{x}\| - \|\hat{y}\| \leq \|\hat{x} + \hat{y}\|.$$

Análogamente, dado que $\|\hat{y}\| = \|(\hat{x} + \hat{y}) + (-\hat{x})\|$, obtenemos que

$$\begin{aligned} \|\hat{y}\| &= \|(\hat{x} + \hat{y}) + (-\hat{x})\| \\ &\leq \|\hat{x} + \hat{y}\| + \|\hat{x}\|, \end{aligned}$$

de donde

$$-\|\hat{x} + \hat{y}\| \leq \|\hat{x}\| - \|\hat{y}\|,$$

es decir, que

$$-\|\hat{x} + \hat{y}\| \leq \|\hat{x}\| - \|\hat{y}\| \leq \|\hat{x} + \hat{y}\|,$$

lo cual es equivalente a la desigualdad que se deseaba probar. ■

Para concluir esta sección, recordemos que la desigualdad de Cauchy-Schwarz nos fue sugerida a partir de la identidad 1.7, identidad que surge de un problema geométrico, el cual a su vez se podía resolver a partir del concepto de ángulo. Todo ello en el plano, o en el conjunto $\mathbb{R}^2$ (pensando a este conjunto como una "representación" de dicho plano).

Lo importante de la prueba anterior es que ésta es válida para cualquier $\mathbb{R}^n$, y con base en la misma identidad, podemos "extender" (y definir de manera más precisa) el concepto de ángulo entre cualquier par de elementos (distintos de $\hat{0}$) de este conjunto, y por lo tanto, definir el concepto de

ángulo entre cualquier par de elementos (distintos de $\hat{0}$) del espacio vectorial representado por el conjunto $\mathbb{R}^n$ (como por ejemplo, el espacio de los polinomios de grado menor o igual a $n-1$). Por todo lo anterior, damos la siguiente

Definición 1.3 *Sean $\hat{x}, \hat{y} \in \mathbb{R}^n$ distintos del $\hat{0}$. Definimos el ángulo (agudo) entre $\hat{x}$ y $\hat{y}$ como el número θ dado por la fórmula*

$$\begin{aligned}\theta &:= \cos^{-1}\left(\frac{\hat{x}\cdot\hat{y}}{\|\hat{x}\|\,\|\hat{y}\|}\right)\\ &:= \arccos\left(\frac{\hat{x}\cdot\hat{y}}{\|\hat{x}\|\,\|\hat{y}\|}\right),\end{aligned}$$

donde $\cos^{-1}$ (o $\arccos$) es la función inversa de $\cos$ que tiene como contradominio el intervalo $[0,\pi]$.

1.4. Otras normas

Como el lector habrá notado, todo lo realizado en la sección anterior está basado en lo que comunmente conocemos como geometría euclideana. La idea de esta sección es mostrar que en particular el concepto de longitud (o magnitud, o distancia al origen, o "valor absoluto") de un elemento $\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ se puede definir de otras formas, sin necesidad de recurrir a la geometría euclideana, y conservando las propiedades elementales que tiene la norma euclideana (y que establecimos en la proposición 1.1).

Sin que se pretenda profundizar en el tema, ahora es prudente mencionar que a cualquier función de $\mathbb{R}^n$ en los reales no negativos (conjunto al que denotaremos por $\mathbb{R}^+$) que satisfaga las propiedades establecidas en la proposición 1.1, se le conoce con el nombre de *norma*. Y aunque existe (literalmente) una infinidad de maneras de definir una norma en $\mathbb{R}^n$ (en realidad, muchas de ellas son muy "parecidas"), en esta breve sección sólo abordaremos dos de ellas; la *norma uno* (que denotaremos por $\|\cdot\|_1$) y la *norma infinito* (que denotaremos por $\|\cdot\|_\infty$), que definimos de la siguiente manera.

Definición 1.4 *Dado $\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, definimos:*

1. *la norma uno de $\hat{x}$, que denotamos por $\|\hat{x}\|_1$, como*

$$\|\hat{x}\|_1 := |x_1| + \cdots + |x_n|$$

2. *la norma infinito de $\hat{x}$, que denotamos por $\|\hat{x}\|_\infty$, como*

$$\|\hat{x}\|_\infty := \text{máx}\{|x_1|, \ldots, |x_n|\}.$$

Queda como un ejercicio para el lector probar que las funciones definidas arriba satisfacen las propiedades enlistadas en la proposición 1.1, con lo que de paso quedará justificado el hecho de que a dichas funciones les llamemos normas.

Estas otras normas tienen algunas diferencias importantes con la norma euclideana. En particular, para ninguna de ellas se puede definir un producto punto en $\mathbb{R}^n$ que satisfaga lo equivalente a la identidad 1.6. Aun así, en cuanto a formas de medir la longitud de un elemento $\hat{x} \in \mathbb{R}^n$, no resultan ser tan diferentes. En efecto, mostraremos que todas estas normas están relacionadas a través de ciertas desigualdades.

Más adelante, con base en el concepto de norma definiremos conceptos tales como: "estar en el interior de un conjunto", "estar en el exterior de un conjunto", y "estar en la frontera de un conjunto". Las desigualdades que se formulan en la siguiente proposición permitirán establecer que todos ellos resultarán ser iguales sin importar qué norma usemos.

Proposición 1.4 *Para cualquier elemento $\hat{x} \in \mathbb{R}^n$ se satisfacen las siguientes desigualdades:*

1. $\|\hat{x}\|_\infty \leq \|\hat{x}\| \leq \sqrt{n}\,\|\hat{x}\|_\infty$
2. $\frac{1}{\sqrt{n}}\,\|\hat{x}\|_1 \leq \|\hat{x}\| \leq \|\hat{x}\|_1$.

Demostración. Dado que el cuadrado de un número real siempre es no negativo, se tiene que

$$x_i^2 \leq x_1^2 + \cdots + x_n^2,$$

de modo que

$$|x_i| \leq \sqrt{x_1^2 + \cdots + x_n^2} = \|\hat{x}\| . \tag{1.9}$$

Como la desigualdad anterior es válida para toda $i \in \{1, \ldots, n\}$, concluimos que

$$\|\hat{x}\|_\infty = \text{máx}\{|x_1| , \ldots , |x_n|\} \leq \|\hat{x}\| .$$

Por otra parte, como

$$|x_i| \leq \text{máx}\{|x_1| , \ldots , |x_n|\} = \|\hat{x}\|_\infty$$

(para toda $i \in \{1, \ldots, n\}$), se tiene que $x_i^2 \leq \|\hat{x}\|_\infty^2$ y por tanto

$$\|\hat{x}\|^2 = x_1^2 + \cdots + x_n^2 \leq \|\hat{x}\|_\infty^2 + \cdots + \|\hat{x}\|_\infty^2 = n \|\hat{x}\|_\infty^2 ,$$

es decir,

$$\|\hat{x}\| \leq \sqrt{n} \|\hat{x}\|_\infty ,$$

con lo que concluimos la prueba de las dos desigualdades del primer inciso.

Para obtener la primera desigualdad del segundo inciso, obsérvese que

$$|x_1| + \cdots + |x_n| = (1, \ldots, 1) \cdot (|x_1| , \ldots , |x_n|),$$

de tal forma que, por la desigualdad de Cauchy-Schwarz, tenemos que

$$\begin{aligned} |x_1| + \cdots + |x_n| &= (1, \ldots, 1) \cdot (|x_1| , \ldots , |x_n|) \\ &\leq \|(1, \ldots, 1)\| \, \|(|x_1| , \ldots , |x_n|)\| \\ &= \sqrt{n} \sqrt{|x_1|^2 + \cdots + |x_n|^2} \\ &= \sqrt{n} \sqrt{x_1^2 + \cdots + x_n^2} \\ &= \sqrt{n} \|\hat{x}\| . \end{aligned}$$

Para obtener la segunda desigualdad, observe que

$$\begin{aligned} \hat{x} &= (x_1, \ldots, x_n) \\ &= (x_1, 0, \ldots, 0) + (0, x_2, 0, \ldots, 0) + \cdots + (0, 0, \ldots, 0, x_n), \end{aligned}$$

de tal forma que, usando el problema 5 de este capítulo, concluimos que

$$\begin{aligned}\|\hat{x}\| &= \|(x_1, 0, \ldots, 0) + (0, x_2, 0, \ldots, 0) + \cdots + (0, 0, \ldots, 0, x_n)\| \\ &\leq \|(x_1, 0, \ldots, 0)\| + \cdots + \|(0, 0, \ldots, 0, x_n)\| \\ &= |x_1| + \cdots + |x_n| \\ &= \|\hat{x}\|_1 ,\end{aligned}$$

que es la desigualdad que se deseaba probar. ■

Para concluir esta sección mencionaremos la importancia adicional que tiene el hecho de que en $\mathbb{R}^n$ podamos contar con al menos una norma (de hecho, hasta ahora en $\mathbb{R}^n$ ¡tenemos tres!).

Como mencionamos anteriormente, el concepto de norma nos permite hablar de la longitud (o magnitud, o distancia al origen, o "valor absoluto") de un elemento $\hat{x} \in \mathbb{R}^n$. Si en particular pensamos a la norma como una forma de medir la distancia que hay entre el "punto" representado por el elemento $\hat{x} \in \mathbb{R}^n$ y el origen $\hat{0}$, esta interpretación nos permite entonces hablar de la distancia entre cualesquiera dos elementos $\hat{x}, \hat{y} \in \mathbb{R}^n$; en efecto, dados estos dos elementos, en virtud de que a $\hat{x} - \hat{y}$ lo podemos pensar como la flecha que tiene punto inicial en $\hat{y}$ y punto final en $\hat{x}$, la norma de este vector será una medida de la distancia entre estos dos puntos.

Si recordamos que en $\mathbb{R}^n$ hemos definido tres normas diferentes, con base en el razonamiento anterior estamos en condiciones de establecer tres formas diferentes de medir la distancia entre elementos de $\mathbb{R}^n$. Sin embargo, y como en el caso de la norma, la distancia que usaremos a lo largo de todo este texto es aquella que se obtiene a partir de la norma euclideana (y que por razones obvias, llamaremos la *distancia euclideana*). Este concepto lo formalizamos en la siguiente

Definición 1.5 *Dados $\hat{x}, \hat{y} \in \mathbb{R}^n$ definimos la distancia (euclideana) de $\hat{x}$ a $\hat{y}$, que denotamos por $d(\hat{x}, \hat{y})$, como*

$$d(\hat{x}, \hat{y}) := \|\hat{x} - \hat{y}\| .$$

Un resultado muy sencillo de probar es que la distancia euclideana (y cualquier otra distancia que se respete), cumple con las propiedades que establecemos en la siguiente proposición[3].

Proposición 1.5 *La distancia (euclideana) satisface las siguientes propiedades:*

1. $d(\hat{x}, \hat{y}) \geq 0$ *para toda* $\hat{x}, \hat{y} \in \mathbb{R}^n$, *y* $d(\hat{x}, \hat{y}) = 0$ *si y sólo si* $\hat{x} = \hat{y}$
2. $d(\hat{x}, \hat{y}) = d(\hat{y}, \hat{x})$ *para toda* $\hat{x}, \hat{y} \in \mathbb{R}^n$
3. $d(\hat{x}, \hat{y}) \leq d(\hat{x}, \hat{z}) + d(\hat{z}, \hat{y})$ *para cualesquiera* $\hat{x}, \hat{y}, \hat{z} \in \mathbb{R}^n$ *(desigualdad del triángulo)*

Como es de esperarse, la prueba de esta proposición se deja al lector.

1.5. Topología de $\mathbb{R}^n$

La Topología es un área de las matemáticas muy importante que se interesa por conceptos como proximidad, continuidad, conectividad (o conexidad), compacidad, y muchos otros más. Para abordarlos de manera precisa, primero es necesario definir un cierto tipo de conjuntos (que en Topología se les conoce como los *conjuntos abiertos*).

Cuando en un conjunto se cuenta con una forma de medir la distancia entre cualesquiera dos de sus elementos (como es el caso de $\mathbb{R}^n$), existe una manera de decir quiénes son los conjuntos abiertos y con base en éstos desarrollar los conceptos que mencionamos al principio de esta sección. Esto es lo que haremos y para ello comenzaremos por definir un cierto tipo de conjunto que resultará ser básico en esta tarea: vecindad de un punto $\hat{x} \in \mathbb{R}^n$.

En $\mathbb{R}^n$, definimos el concepto de vecindad apoyándonos en alguna de las formas que disponemos para medir la distancia entre elementos de $\mathbb{R}^n$, que

[3]De hecho, a cualquier función de $\mathbb{R}^n \times \mathbb{R}^n$ en $\mathbb{R}^+$ que cumpla con estas tres propiedades se le llamará *una distancia* en $\mathbb{R}^n$. Aún y cuando en este texto podemos considerar tres diferentes distancias, con base en las tres diferentes normas que definimos, no todas las distancias en $\mathbb{R}^n$ tienen que estar definidas a través de una norma.

en nuestro caso será la distancia euclideana (en los problemas mostraremos que, en términos topológicos, da lo mismo cuál de estas distancias elijamos o, en última instancia, cuál norma elijamos).

Algunos objetos geométricos son muy sencillos de describir en términos del concepto de distancia, y sin duda el más sencillo de ellos es el formado por los puntos que se encuentran a una distancia constante $r > 0$ de un punto fijo $\hat{x} \in \mathbb{R}^n$, que como todos sabemos, en $\mathbb{R}^2$ es una circunferencia de radio r, y en $\mathbb{R}^3$ es una esfera, también de radio r. Sin embargo, para nuestros objetivos, el conjunto que resultará de mayor interés no es tanto el formado por los puntos que se encuentran a una distacia r de $\hat{x} \in \mathbb{R}^n$, sino los que se encuentran a una distancia *menor a* r. En $\mathbb{R}^2$ este conjunto consistirá de los puntos que se encuentran "dentro" de la cincunferencia de radio r, y en $\mathbb{R}^3$ consistirá de los que están "dentro" de la esfera, también de radio r (ver figura 1.9).

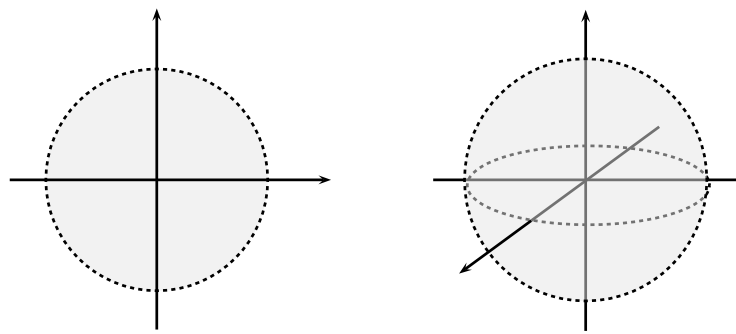


Figura 1.9: Las bolas de radio r con centro en el origen (en $\mathbb{R}^2$ y en $\mathbb{R}^3$).

Con base en lo anterior, definimos (en $\mathbb{R}^n$) el concepto de *vecindad* (*o bola*) *de radio* $r > 0$ *con centro en el punto* $\hat{x} \in \mathbb{R}^n$, que denotaremos por $B_r(\hat{x})$, y formulamos en la siguiente

Definición 1.6 *Dado* $\hat{x} \in \mathbb{R}^n$ *y* $r > 0$*, definimos la vecindad (o bola) de radio* $r > 0$ *con centro en* $\hat{x} \in \mathbb{R}^n$ *como*

$$B_r(\hat{x}) := \{\hat{y} \in \mathbb{R}^n \mid d(\hat{x}, \hat{y}) = \|\hat{x} - \hat{y}\| < r\}.$$

1.5.1. Clasificación de puntos

Apoyándonos en este concepto, ahora nos daremos a la tarea de, dado un conjunto $A \subset \mathbb{R}^n$, clasificar a todos los puntos de $\mathbb{R}^n$ en términos de su "localización" con respecto a dicho conjunto, en donde por localización nos referimos a algo más profundo que el simple concepto de pertenencia. A fin de precisar esta idea de localización con respecto a un conjunto A, recurriremos a la figura 1.10 en $\mathbb{R}^2$.

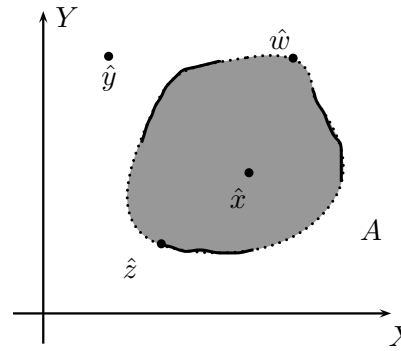


Figura 1.10: Las diferentes localizaciones que puede tener un punto de $\mathbb{R}^n$ con respecto a un conjunto A.

Suponiendo que el conjunto A está formado por los puntos del área sombreada, incluyendo los puntos de la línea continua y excluyendo los puntos de la línea punteada, tenemos que los puntos $\hat{x}$ y $\hat{z}$ pertenecen al conjunto, mientras que los puntos $\hat{y}$ y $\hat{w}$ no pertenecen a A. Sin embargo, tanto los que pertenecen como los que no pertenecen a A, tienen una localización diferente con respecto a este conjunto.

En efecto, mientras que en el caso del punto $\hat{x}$ *existe un radio* $r > 0$ tal que la vecindad de este radio y centro en $\hat{x}$ está contenida en A (es decir, $B_r(\hat{x}) \subset A$), en el caso del punto $\hat{z}$ no es posible encontrar un radio con esta misma característica; es decir, *para todo* $r > 0$ se tiene que $B_r(\hat{z}) \cap A^c \neq \varnothing$ y $B_r(\hat{z}) \cap A \neq \varnothing$, en donde $A^c := \mathbb{R}^n \setminus A$ (el complemento de A).

En el caso de los puntos $\hat{y}$ y $\hat{w}$ se tiene una situación análoga; mientras que para el punto $\hat{y}$ *existe un radio* $r' > 0$ tal que la vecindad de este radio y centro en $\hat{y}$ está contenida en A^c (es decir, $B_{r'}(\hat{y}) \subset A^c$), en el caso del punto $\hat{w}$ también se tiene que *para todo* $r > 0$, sucede que $B_r(\hat{w}) \cap A^c \neq \varnothing$ y $B_r(\hat{w}) \cap A \neq \varnothing$ (veáse la figura 1.11).

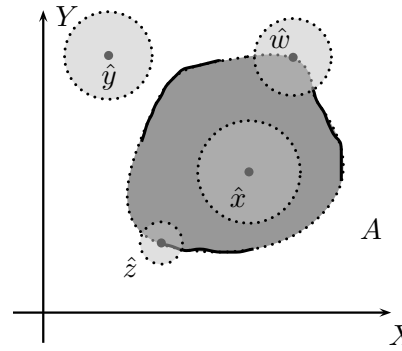


Figura 1.11: Caracterización, en términos de vecindades, de la localización de los puntos de $\mathbb{R}^n$ con respecto al conjunto A.

Resumiendo lo anterior, tenemos que, si $A \subset \mathbb{R}^n$ y $\hat{x} \in \mathbb{R}^n$, entonces se satisface una de las siguientes condiciones (y sólo una de ellas, pues serán mutuamente excluyentes):

1. existe $r > 0$ tal que $B_r(\hat{x}) \subset A$,
2. existe $r > 0$ tal que $B_r(\hat{x}) \subset A^c$, o
3. para todo $r > 0$ se cumple que $B_r(\hat{x}) \cap A^c \neq \varnothing$ y $B_r(\hat{x}) \cap A \neq \varnothing$.

Seguramente el lector estará de acuerdo en que los puntos que satisfacen la condición 1 no sólo son puntos que deben pertenecer a A, sino que además están realmente "dentro" (o en "el interior") de A; los que cumplen la condición 2 no sólo son puntos que no pertenecen a A, sino que además están realmente "fuera" (o en "el exterior") de A; y finalmente, los que cumplen la condición 3 pueden o no pertenecer a A, pero lo importante es que no están

ni "dentro" ni "fuera", razón por la cual podemos decir que se encuentran en el "borde" (o en "la frontera") de A.

Con base en lo anterior es que ahora estamos en condiciones de, dado un conjunto $A \subset \mathbb{R}^n$, dar una clasificación de los puntos de $\mathbb{R}^n$ en términos de su localización con respecto a dicho conjunto, cosa que haremos en la siguiente

Definición 1.7 *Sean $A \subset \mathbb{R}^n$ y $\hat{x} \in \mathbb{R}^n$. Decimos que:*

1. *$\hat{x}$ es un punto interior de A si existe $r > 0$ tal que $B_r(\hat{x}) \subset A$. Denotamos por $\operatorname{int}(A)$ al conjunto formado por todos estos puntos, es decir*
$$\operatorname{int}(A) := \{\hat{x} \in \mathbb{R}^n \mid \hat{x} \text{ es un punto interior de } A\}$$
y diremos que este conjunto es el interior de A.

2. *$\hat{x}$ es un punto exterior de A si existe $r > 0$ tal que $B_r(\hat{x}) \subset A^c$. Denotamos por $\operatorname{ext}(A)$ al conjunto formado por todos estos puntos, es decir*
$$\operatorname{ext}(A) := \{\hat{x} \in \mathbb{R}^n \mid \hat{x} \text{ es un punto exterior de } A\}$$
y diremos que este conjunto es el exterior de A.

3. *$\hat{x}$ es un punto frontera de A si para todo $r > 0$ se tiene que $B_r(\hat{x}) \cap A \neq \varnothing$ y $B_r(\hat{x}) \cap A^c \neq \varnothing$. Denotamos por $\operatorname{Fr}(A)$ al conjunto formado por todos estos puntos, es decir*
$$\operatorname{Fr}(A) := \{\hat{x} \in \mathbb{R}^n \mid \hat{x} \text{ es un punto frontera de } A\}$$
y diremos que este conjunto es la frontera de A.

Como seguramente el lector ya habrá notado (y podrá probar muy fácilmente), los conjuntos definidos anteriormente satisfacen unas propiedades muy elementales (además de algunas otras que veremos más adelante) y que dejamos expresadas en la siguiente

Proposición 1.6 *Si $A \subset \mathbb{R}^n$, entonces:*

1. $\text{int}(A) \subset A$
2. $\text{ext}(A) \subset A^c$
3. $\text{int}(A) \cap \text{ext}(A) = \text{int}(A) \cap \text{Fr}(A) = \text{Fr}(A) \cap \text{ext}(A) = \varnothing$
4. $\mathbb{R}^n = \text{int}(A) \cup \text{Fr}(A) \cup \text{ext}(A)$
5. $\text{int}(A^c) = \text{ext}(A)$ *y* $\text{Fr}(A) = \text{Fr}(A^c)$.

Aun cuando la definición del tipo de puntos (y sus repectivos conjuntos) que acabamos de dar fue motivada a partir de un subconjunto de $\mathbb{R}^2$ muy "bonito", hay casos para los cuales dichos conjuntos no resultan ser los que esperamos que sean (¡claro, si acaso tuviéramos una idea de quiénes deberían de ser!). El ejemplo que a continuación presentamos es uno de esos casos.

Ejemplo 1.7 *Sea*

$$A = ([0,1]\times[0,1])\cap(\mathbb{Q}\times\mathbb{Q}) = \{(x,y) \in \mathbb{R}^2 \mid x, y \in \mathbb{Q} \text{ y } 0 \leq x \leq 1, 0 \leq y \leq 1\}.$$

Determinaremos quiénes son el $\text{int}(A)$*, el* $\text{ext}(A)$ *y la* $\text{Fr}(A)$.

Primero mostraremos que si $\hat{x} = (x,y)$ es un punto arbitrario de $\mathbb{R}^2$ y r es cualquier número real positivo, entonces $B_r(\hat{x}) \cap (\mathbb{R}^2 \setminus \mathbb{Q} \times \mathbb{Q}) \neq \varnothing$ de tal forma que, como $A \subset \mathbb{Q}\times\mathbb{Q}$, tendremos que $B_r(\hat{x})\cap(\mathbb{R}^2 \setminus A) = B_r(\hat{x})\cap A^c \neq \varnothing$. Sea pues $\hat{x} = (x,y) \in \mathbb{R}^2$ y r cualquier número real positivo.

Como es por todos conocido, existe $x' \notin \mathbb{Q}$ tal que $x < x' < x + r$. De estas desigualdades concluimos que si $\hat{y} = (x', y)$, entonces $d(\hat{x}, \hat{y}) = \|(x,y) - (x',y)\| = |x - x'| < r$ de modo que $\hat{y} \in B_r(\hat{x})$, y como $\hat{y} \notin \mathbb{Q} \times \mathbb{Q}$ (pues $x' \notin \mathbb{Q}$), entonces $B_r(\hat{x}) \cap (\mathbb{R}^2 \setminus \mathbb{Q} \times \mathbb{Q}) \neq \varnothing$.

De lo que acabamos de probar se desprende inmediatamente que $\text{int}(A) = \varnothing$ *pues si $\hat{x} \in A$, no existe forma de encontrar $r > 0$ tal que $B_r(\hat{x}) \subset A$, pues cualquier vecindad, sin importar cuál es su radio (e incluso sin importar cuál es su centro) interseca a A^c de tal forma que por el inciso* 1 *de la proposición 1.6 tendremos que* $\text{int}(A) = \varnothing$.

El siguiente paso será mostrar que $[0,1]\times[0,1] \subset \mathrm{Fr}(A)$. *Sea* $\hat{x} = (x,y) \in [0,1]\times[0,1]$ *y* $r > 0$. *Por el primer resultado que probamos, ya sabemos que* $B_r(\hat{x}) \cap A^c \neq \varnothing$. *Restaría probar que* $B_r(\hat{x}) \cap A \neq \varnothing$.

Supongamos por ahora que $0 \leq x < 1$ *y* $0 \leq y < 1$. *Por la densidad de los números racionales sabemos que existen* $x', y' \in \mathbb{Q}$ *tales que* $x < x' < \min\{1, x + r/\sqrt{2}\}$ *y* $y < y' < \min\{1, y + r/\sqrt{2}\}$. *Tenemos entonces que* $(x', y') \in A$ *y además, dado que* $|x - x'| < r/\sqrt{2}$ *y* $|y - y'| < r/\sqrt{2}$, *entonces*

$$\begin{aligned} \big\|(x,y) - (x',y')\big\| &= \sqrt{(x-x')^2 + (y-y')^2} \\ &< \sqrt{(r/\sqrt{2})^2 + (r/\sqrt{2})^2} \\ &= r \end{aligned}$$

de modo que $(x', y') \in B_r(\hat{x})$, *es decir,* $B_r(\hat{x}) \cap A \neq \varnothing$.

Ahora, escogiendo a x' *y* y' *como en el caso anterior (siempre que se pueda), si* $x = 1$ *y* $y < 1$, *entonces nos fijamos en la pareja* $(1, y')$; *si* $x < 1$ *y* $y = 1$, *elegimos a la pareja* $(x', 1)$; *finalmente, si* $x = 1$ *y* $y = 1$, *elegimos a la pareja* $(1,1)$. *En todos estos casos, dichas parejas pertenecen a* $B_r(\hat{x}) \cap A$, *con lo cual queda probado también en estos casos que* $B_r(\hat{x}) \cap A \neq \varnothing$, *y por lo tanto que* $[0,1] \times [0,1] \subset \mathrm{Fr}(A)$.

Finalmente, probaremos que $\mathbb{R}^2 \setminus [0,1] \times [0,1] \subset \mathrm{ext}(A)$. *Sea* $\hat{x} = (x,y) \notin [0,1] \times [0,1]$ *y supongamos por ahora que* $x < 0$ *o* $1 < x$ *(la otra posibilidad es que* $y < 0$ *o* $1 < y$*). Si* $x < 0$ *tomamos* $r = |x| > 0$ *y afirmamos que* $B_r(\hat{x}) \subset \mathbb{R}^2 \setminus [0,1] \times [0,1] \subset A^c$. *En efecto, si* $(x', y') \in B_r(\hat{x})$, *por la desigualdad 1.9 se tiene que*

$$\begin{aligned} \big|x - x'\big| &\leq \big\|(x,y) - (x',y')\big\| \\ &< r \\ &= |x| \\ &= -x. \end{aligned}$$

De esta forma, se tiene que $x < x - x' < -x$ *y de la primera desigualdad concluimos que* $x' < 0$. *Por lo tanto,* $(x', y') \notin [0,1] \times [0,1]$, *con lo que probamos que* $B_r(\hat{x}) \subset \mathbb{R}^2 \setminus [0,1] \times [0,1] \subset A^c$. *Si* $1 < x$ *tomamos* $r = x - 1 > 0$

y afirmamos nuevamente que $B_r(\hat{x}) \subset \mathbb{R}^2 \setminus [0,1] \times [0,1] \subset A^c$. *En efecto, si* $(x', y') \in B_r(\hat{x})$, *otra vez por la desigualdad 1.9 se tiene que*

$$\begin{aligned} |x - x'| &\leq \|(x, y) - (x', y')\| \\ &< r \\ &= x - 1, \end{aligned}$$

de modo que $1 - x < x - x' < x - 1$. *Por la segunda desigualdad, concluimos que* $1 < x'$ *y por lo tanto se tiene que* $(x', y') \notin [0,1] \times [0,1]$, *con lo que probamos también que* $B_r(\hat{x}) \subset \mathbb{R}^2 \setminus [0,1] \times [0,1] \subset A^c$. *Si* $y < 0$ *o* $1 < y$ *procedemos de forma análoga.*

Reuniendo lo que hemos probado hasta ahora, por los incisos 3 *y* 4 *de la proposición 1.6, tenemos que*

$$[0,1] \times [0,1] \subset \mathrm{Fr}(A) = \mathbb{R}^n \setminus \mathrm{ext}(A) \subset [0,1] \times [0,1],$$

de donde concluimos que $\mathrm{Fr}(A) = [0,1] \times [0,1]$ *y que* $\mathrm{ext}(A) = \mathbb{R}^n \setminus [0,1] \times [0,1]$.

1.5.2. Conjuntos abiertos y cerrados

Una vez que hemos clasificado los puntos de $\mathbb{R}^n$ con respecto a un conjunto $A \subset \mathbb{R}^n$, estamos en condiciones de dar una caracterización de cierto tipo de subconjuntos de $\mathbb{R}^n$.

Dado un conjunto $A \subset \mathbb{R}^n$, ya sabemos que $\mathrm{int}(A) \subset A$ y que $\mathrm{ext}(A) \subset A^c$. Sin embargo, en el caso de la $\mathrm{Fr}(A)$ no siempre se cumple alguna de estas relaciones de contención. El conjunto A de la figura 1.11 ilustra un caso de este tipo. Lo que sí se cumple en general es que $A = \mathrm{int}(A) \cup (A \cap \mathrm{Fr}(A))$, identidad que siempre vale la pena recordar.

Los conjuntos para los cuales se cumple que $\mathrm{Fr}(A) \subset A$ o $\mathrm{Fr}(A) \subset A^c$ (esto último equivalente a que $\mathrm{Fr}(A) \cap A = \varnothing$) merecen un nombre aparte. Cuando un conjunto no tenga a ninguno de sus puntos frontera (es decir que $\mathrm{Fr}(A) \cap A = \varnothing$), diremos que el conjunto es *abierto*, y si $\mathrm{Fr}(A) \subset A$, diremos que el conjunto es *cerrado*. Dejamos expresada esta caracterización de conjuntos en la siguiente

Definición 1.8 *Sea $A \subset \mathbb{R}^n$. Decimos que:*

1. *A es un conjunto abierto si $\mathrm{Fr}(A) \cap A = \varnothing$.*
2. *A es un conjunto cerrado si $\mathrm{Fr}(A) \subset A$.*

Una primera observación que es muy importante hacer con relación a las definiciones anteriores es que las condiciones $\mathrm{Fr}(A) \cap A = \varnothing$ y $\mathrm{Fr}(A) \subset A$ no son excluyentes: la negación de una de ellas no es la otra. Es decir, el que $\mathrm{Fr}(A) \cap A \neq \varnothing$ no implica que $\mathrm{Fr}(A) \subset A$, y recíprocamente, el que $\mathrm{Fr}(A) \nsubseteq A$ no implica que $\mathrm{Fr}(A) \cap A = \varnothing$. Dicho de otra forma, en términos de las definiciones que acabamos de dar: *que un conjunto A no sea abierto no implica que tenga que ser cerrado*, y recíprocamente, *que un conjunto A no sea cerrado no implica que tenga que ser abierto.* Más aún, dado que las condiciones $\mathrm{Fr}(A) \cap A \neq \varnothing$ y $\mathrm{Fr}(A) \nsubseteq A$ no se contraponen (es decir, pueden pasar ambas, como sucede con el conjunto A de la figura 1.11), confirmamos que *hay conjuntos que no son abiertos y tampoco son cerrados*, y que son justo aquellos que sólo contienen a algunos de sus puntos frontera, pero no a todos ellos.

La segunda observación importante con relación a las definiciones que acabamos de dar es que éstas no son las más frecuentes, razón por la cual lo primero que haremos será probar una proposición en la que se establecen las equivalencias con las definiciones más "populares" (o más conocidas, si se prefiere).

Proposición 1.7 *Sea $A \subset \mathbb{R}^n$. Las siguientes afirmaciones son equivalentes:*

1. *A es un conjunto abierto si y sólo si para todo $\hat{x} \in A$ existe $r > 0$ tal que $B_r(\hat{x}) \subset A$ (es decir, si $A \subset \mathrm{int}(A)$ y por tanto que $A = \mathrm{int}(A)$).*
2. *A es un conjunto cerrado si y sólo si A^c es abierto.*

Demostración. Inciso 1)
$\Rightarrow$]
Sea $\hat{x} \in A$; de acuerdo con el inciso 1 de la definición 1.8 se tiene que

$\hat{x} \notin \text{Fr}(A)$ y por lo tanto, de acuerdo con el inciso 3 de la definición 1.7, debe existir $r > 0$ tal que $B_r(\hat{x}) \cap A = \varnothing$ o $B_r(\hat{x}) \cap A^c = \varnothing$. Como $\hat{x} \in B_r(\hat{x}) \cap A$, la identidad que se debe cumplir es la segunda, es decir, que $B_r(\hat{x}) \cap A^c = \varnothing$, lo que significa que $B_r(\hat{x}) \subset A$ (y por lo tanto que $\hat{x} \in \text{int}(A)$).

Inciso 1)

$\Leftarrow]$

Dado que $A = A \cap \mathbb{R}^n$, por los incisos 2, 3 y 4 de la proposición 1.6 se tiene que

$$A = A \cap \mathbb{R}^n = (A \cap \text{int}(A)) \cup (A \cap \text{Fr}(A)) \cup (A \cap \text{ext}(A)) = A \cap \text{int}(A),$$

de donde $A \subset \text{int}(A)$ (y por el inciso 1 de la misma proposición, se tiene que $A = \text{int}(A)$).

Inciso 2)

$\Rightarrow]$

De acuerdo con el inciso 1 de la definición 1.8, como A es cerrado, se tiene que $\text{Fr}(A) \subset A$ y por tanto que $\text{Fr}(A) \cap A^c = \varnothing$; dado que $\text{Fr}(A) = \text{Fr}(A^c)$ (inciso 5 de la proposición 1.6), tenemos que $\text{Fr}(A^c) \cap A^c = \varnothing$ y por lo tanto, de acuerdo con la definición 1.8, se tiene que A^c es un conjunto abierto.

Inciso 2)

$\Leftarrow]$

¡Lea "hacia atrás" el párrafo anterior! ■

En realidad, y para ser sinceros, las condiciones de la proposición anterior son las más populares porque suelen ser las más útiles cuando se trata de demostrar que un conjunto es abierto o que un conjunto es cerrado. Como prueba de esta afirmación, a continuación damos algunos ejemplos.

Ejemplo 1.8 *Observe que:*

1. $\mathbb{R}^n$ *es un conjunto abierto. En efecto, nótese que para todo* $\hat{x} \in \mathbb{R}^n$, *se tiene que* $B_r(\hat{x}) \subset \mathbb{R}^n$ *sin importar cómo sea* r *(muy grande o muy pequeño). De forma, por el inciso* 1 *de la proposición anterior,* $\mathbb{R}^n$ *es un conjunto abierto.*

2. *$\mathbb{R}^n$ es un conjunto cerrado. En virtud de lo anterior, tenemos que $\mathbb{R}^n = \text{int}(\mathbb{R}^n)$, de forma tal que $\text{Fr}(\mathbb{R}^n) = \varnothing$ (de hecho, también $\text{ext}(\mathbb{R}^n) = \varnothing$). De esta modo, $\text{Fr}(\mathbb{R}^n) \subset \mathbb{R}^n$ y por el inciso 2 de la proposición anterior, $\mathbb{R}^n$ también es un conjunto cerrado.*

3. *Por los dos incisos anteriores (y la multicitada proposición), se concluye que el conjunto $\varnothing$ también es un conjunto abierto y cerrado (lo que por cierto prueba que los conjuntos (al menos algunos de ellos) no son como las puertas: los hay algunos que son ¡abiertos y cerrados al mismo tiempo! ¿El lector podría mostrar otro ejemplo con la misma característica?).*

4. *Para todo $\hat{x} \in \mathbb{R}^n$ y toda $r > 0$ se tiene que $B_r(\hat{x})$ es un conjunto abierto. Para probarlo, veremos que se cumple la condición del inciso 1 de la susodicha proposición. Sea $\hat{y} \in B_r(\hat{x})$; como se ve en la figura 1.12 (para el caso de $\mathbb{R}^2$), parece que si tomamos $r' = r - \|\hat{x} - \hat{y}\| > 0$, entonces se debe tener que $B_{r'}(\hat{y}) \subset B_r(\hat{x})$. La clave para probar esta contención está en la desigualdad del triángulo (inciso 1 de la proposición 1.3) pues, con base en dicha desigualdad, tenemos que si $\hat{z} \in B_{r'}(\hat{y})$, entonces*

$$\begin{aligned} \|\hat{x} - \hat{z}\| &= \|(\hat{x} - \hat{y}) + (\hat{y} - \hat{z})\| \\ &\leq \|\hat{x} - \hat{y}\| + \|\hat{y} - \hat{z}\| \\ &= \|\hat{x} - \hat{y}\| + \|\hat{z} - \hat{y}\| \\ &< \|\hat{x} - \hat{y}\| + r - \|\hat{x} - \hat{y}\| \\ &= r \end{aligned}$$

y por lo tanto $\hat{z} \in B_r(\hat{x})$, de donde $B_{r'}(\hat{y}) \subset B_r(\hat{x})$. Esto prueba que se satisface la condición del inciso 1 de la famosa proposición, y por lo tanto que $B_r(\hat{x})$ es un conjunto abierto.

Pareciera geométricamente "natural" esperar que si a un conjunto A le "quitamos" sus puntos frontera, el conjunto que nos queda será un conjunto abierto (el conjunto de la figura 1.12 refuerza esta sospecha). Cuando a

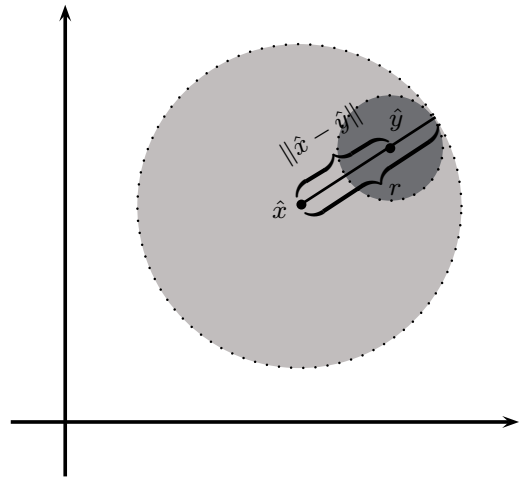


Figura 1.12: La bola de radio r con centro en $\hat{x}$ es un conjunto abierto.

un conjunto A le quitamos sus puntos frontera, sólo nos quedamos con los puntos interiores de A, es decir que

$$A \setminus \mathrm{Fr}(A) = \mathrm{int}(A), \tag{1.10}$$

de donde nuestra sospecha se traduce en que el interior de todo subconjunto A de $\mathbb{R}^n$ ($\mathrm{int}(A)$) será un conjunto abierto (y por lo tanto el $\mathrm{ext}(A)$ también será abierto dado que $\mathrm{ext}(A) = \mathrm{int}(A^c)$ (inciso 5 de la proposición 1.6)).

Siguiendo la misma línea de pensamiento que en el párrafo anterior, ahora nos podemos preguntar qué sucede si a un conjunto A, en lugar de "quitarle" sus puntos frontera, se los agregamos. Es decir, ¿qué tipo de conjunto resultará ser $A \cup \mathrm{Fr}(A)$?

Pues bien, como seguramente el lector ya intuía, este conjunto siempre resultará ser un conjunto cerrado y hasta tiene un nombre que refleja esta propiedad: *la cerradura* de A. Usaremos mucho este conjunto a lo largo de este texto, y por esta razón le dedicamos la siguiente definición.

Definición 1.9 *Sea $A \subset \mathbb{R}^n$. Definimos la cerradura de A, que denotamos por $\bar{A}$, como*

$$\bar{A} := A \cup \mathrm{Fr}(A),$$

Resumiendo lo hecho hasta aquí, dado $A \subset \mathbb{R}^n$ arbitrario, hemos definido (por ahora) cuatro conjuntos asociados a dicho conjunto: $\mathrm{int}(A)$, $\mathrm{ext}(A)$, $\mathrm{Fr}(A)$ y $\bar{A}$. Ahora, si recordamos que, por el inciso 4 de la proposición 1.6, se tiene que

$$(\mathrm{Fr}(A))^c = \mathbb{R}^n \setminus \mathrm{Fr}(A) = \mathrm{int}(A) \cup \mathrm{ext}(A),$$

todo parece indicar que el complemento de la $\mathrm{Fr}(A)$ será un conjunto abierto, y por lo tanto éste será cerrado. Es decir, de estos cuatro conjuntos, el $\mathrm{int}(A)$ y el $\mathrm{ext}(A)$ serán conjuntos abiertos, mientras que $\mathrm{Fr}(A)$ y $\bar{A}$ serán conjuntos cerrados. Dada la importancia de estas propiedades, las dejaremos establecidas en la siguiente

Proposición 1.8 *Sea $A \subset \mathbb{R}^n$ arbitrario. Las siguientes afirmaciones son ciertas:*

1. *el* $\mathrm{int}(A)$ *y el* $\mathrm{ext}(A)$ *son conjuntos abiertos,*
2. *la* $\mathrm{Fr}(A)$ *y* $\bar{A}$ *son conjuntos cerrados.*

Demostración. Inciso 1) De acuerdo con el inciso 1 de la proposición 1.7, necesitamos mostrar que si $\hat{x} \in \mathrm{int}(A)$, entonces existe $r > 0$ tal que $B_r(\hat{x}) \subset \mathrm{int}(A)$. Ahora, como $\hat{x} \in \mathrm{int}(A)$, sabemos que existe $r > 0$ tal que $B_r(\hat{x}) \subset A$; aseguramos que $B_r(\hat{x}) \subset \mathrm{int}(A)$. En efecto, si $\hat{y} \in B_r(\hat{x})$, como se mostró en el inciso 4 del ejemplo 1.8, existe $r' > 0$ tal que $B_{r'}(\hat{y}) \subset B_r(\hat{x}) \subset A$, lo que prueba que $\hat{y} \in \mathrm{int}(A)$ y por tanto que $B_r(\hat{x}) \subset \mathrm{int}(A)$.

En cuanto al $\mathrm{ext}(A)$, como mencionamos arriba, dado que $\mathrm{ext}(A) = \mathrm{int}(A^c)$ (inciso 5 de la proposición 1.6), por la primera parte de este inciso tenemos que este conjunto también es abierto.

Inciso 2) Como también hicimos notar anteriormente, por el inciso 4 de la proposición 1.6, tenemos que

$$(\mathrm{Fr}(A))^c = \mathbb{R}^n \setminus \mathrm{Fr}(A) = \mathrm{int}(A) \cup \mathrm{ext}(A),$$

de tal forma que, por el inciso (a) del problema 21 concluimos que $(\mathrm{Fr}(A))^c$ es un conjunto abierto y por lo tanto que $\mathrm{Fr}(A)$ es cerrado. Con respecto a $\bar{A}$ obsérvese que, como $A = \mathrm{int}(A) \cup (A \cap \mathrm{Fr}(A))$, entonces $A \cup \mathrm{Fr}(A) = \mathrm{int}(A) \cup \mathrm{Fr}(A)$ y por tanto

$$(\bar{A})^c = \mathbb{R}^n \setminus (A \cup \mathrm{Fr}(A)) = \mathbb{R}^n \setminus (\mathrm{int}(A) \cup \mathrm{Fr}(A)) = \mathrm{ext}(A),$$

de modo que $(\bar{A})^c$ es un conjunto abierto y por lo tanto $\bar{A}$ será un conjunto cerrado. ■

1.5.3. Otra clasificación de puntos

Si en la definición 1.7 hicimos una clasificación de los puntos de $\mathbb{R}^n$ en términos de su localización con respecto a un conjunto $A \subset \mathbb{R}^n$ (más allá de su relación de pertenencia con dicho conjunto), ahora introduciremos una nueva clasificación en la que trataremos de reflejar si un punto $\hat{x} \in \mathbb{R}^n$ está "pegado a" (o contrariamente "aislado de") un conjunto $A \subset \mathbb{R}^n$. Para hacer esto, nuevamente haremos uso de las vencidades (o bolas) centradas en el punto $\hat{x}$.

Parece "natural" pensar que un punto $\hat{x} \in \mathbb{R}^n$, el cual ni siquiera tiene que pertenecer a A, está "pegado" a un conjunto $A \subset \mathbb{R}^n$ si cualquiera de sus vecindades $B_r(\hat{x})$ (es decir, sin importar el tamao de r) "comparte" muchos puntos con A.

Con base en esta idea intuitiva es que definimos los conceptos de *punto de acumulación* y *punto aislado* de un conjunto $A \subset \mathbb{R}^n$.

Definición 1.10 *Sean $A \subset \mathbb{R}^n$ y $\hat{x} \in \mathbb{R}^n$. Decimos que:*

1. *$\hat{x}$ es un punto de acumulación de A si para toda $r > 0$ se tiene que $(B_r(\hat{x}) \setminus \{\hat{x}\}) \cap A \neq \varnothing$[4]. Al conjunto formado por los puntos de acumulación de A lo denotamos por A', es decir*

$$A' := \{\hat{x} \in \mathbb{R}^n \mid \hat{x} \text{ es punto de acumulación de } A\},$$

[4]Se dice que de la forma $B_r(\hat{x}) \setminus \{\hat{x}\}$ es una vecindad "agujerada" de $\hat{x}$.

2. *$\hat{x} \in A$ es un punto aislado de A si $\hat{x}$ no es un punto de acumulación de A, es decir, si existe $r > 0$ tal que $(B_r(\hat{x}) \setminus \{\hat{x}\}) \cap A = \varnothing$.*

Es muy importante destacar que la relación definida en el inciso 1 entre un punto $\hat{x}$ y un conjunto A es completamente independiente de la relación de pertenencia que $\hat{x}$ tenga con el conjunto A. En efecto, si $\hat{x}$ es un punto de acumulación de A, esto no significa que $\hat{x}$ tenga que pertenecer a A, y recíprocamente, aún cuando $\hat{x} \in A$, puede suceder que $\hat{x}$ no sea un punto de acumulación de A, es decir, puede suceder que $\hat{x}$ sea un punto aislado de A. El siguiente ejemplo ilustra (en $\mathbb{R}^2$) estos hechos.

Ejemplo 1.9 *Sea $A = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\} \cup \{(2,2)\} = B_1((0,0)) \cup \{(2,2)\}$ (ver figura 1.13). Afirmamos que:*

1. *el punto $(1/\sqrt{2}, 1/\sqrt{2})$, que no pertenece a A, es un punto de acumulación de A. En efecto, observe que dado $r > 0$, se tiene que el punto*

$$\left(\frac{1}{\sqrt{2}} - \frac{r}{\sqrt{2}(r+1)}, \frac{1}{\sqrt{2}} - \frac{r}{\sqrt{2}(r+1)}\right) = \frac{1}{\sqrt{2}(r+1)}(1,1)$$

es tal que

$$\begin{aligned}\left\|\left(\frac{1}{\sqrt{2}} - \frac{r}{\sqrt{2}(r+1)}, \frac{1}{\sqrt{2}} - \frac{r}{\sqrt{2}(r+1)}\right)\right\| &= \frac{1}{\sqrt{2}(r+1)}\|(1,1)\| \\ &= \frac{1}{\sqrt{2}(r+1)}\sqrt{2} \\ &= \frac{1}{r+1} \\ &< 1\end{aligned}$$

y por lo tanto pertenece a A.

Por otra parte, se tiene que

$$\begin{aligned}
0 &< \left\| \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) - \left(\frac{1}{\sqrt{2}} - \frac{r}{\sqrt{2}(r+1)}, \frac{1}{\sqrt{2}} - \frac{r}{\sqrt{2}(r+1)} \right) \right\| \\
&= \left\| \left(\frac{r}{\sqrt{2}(r+1)}, \frac{r}{\sqrt{2}(r+1)} \right) \right\| \\
&= \frac{r}{\sqrt{2}(r+1)} \|(1,1)\| \\
&= \frac{r}{\sqrt{2}(r+1)} \sqrt{2} \\
&= \frac{r}{(r+1)} \\
&< r,
\end{aligned}$$

de donde concluimos que este punto también pertenece al conjunto

$$B_r((1/\sqrt{2}, 1/\sqrt{2})) \setminus \{(1/\sqrt{2}, 1/\sqrt{2})\}$$

y por lo tanto que

$$\left(B_r((1/\sqrt{2}, 1/\sqrt{2})) \setminus \{(1/\sqrt{2}, 1/\sqrt{2})\} \right) \cap A \neq \varnothing,$$

es decir, que $(1/\sqrt{2}, 1/\sqrt{2})$ *es un punto de acumulación de* A*.*

2. *el punto* $(2,2)$*, que es un punto que pertenece a* A*, es un punto aislado de* A*. Obsérvese que si tomamos la bola (o vecindad) de radio* 1 *con centro en este punto, entonces*

$$(B_1((2,2)) \setminus \{(2,2)\}) \cap A = \varnothing.$$

En efecto, si $(x,y) \in B_1((2,2)) \setminus \{(2,2)\}$*, entonces* $(x,y) \neq (2,2)$*. Además se cumplen las siguientes desigualdades:*

$$\|(2,2)\| - \|(x,y)\| \leq \|(2,2) - (x,y)\| < 1,$$

de modo que

$$\|(2,2)\| - 1 < \|(x,y)\|.$$

Como

$$1 < 2\sqrt{2} - 1 = \|(2,2)\| - 1 < \|(x,y)\|\,,$$

entonces $1 < x^2 + y^2$*, es decir,* $(x,y) \notin B_1((2,2))$ *y por lo tanto* $(x,y) \notin A$*.*

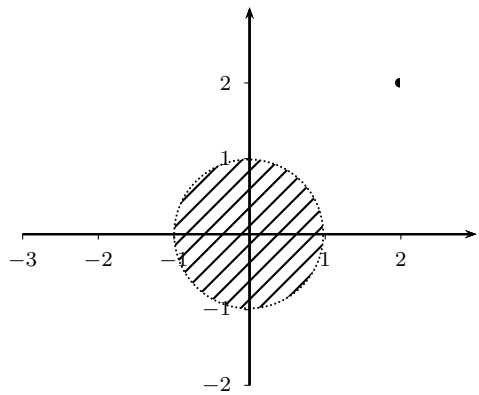


Figura 1.13: El conjunto $A = \{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 < 1\} \cup \{(2,2)\}$ del ejemplo 1.9.

Dado un conjunto $A \subset \mathbb{R}^n$, la relación de sus puntos de acumulación con aquellos que se definieron en 1.7 no es muy estrecha, pues cada uno de ellos expresa una relación diferente con dicho conjunto. Sin embargo, existen algunas relaciones entre ellos que es importante destacar. Tal es el caso de los puntos interiores de A, los cuales siempre resultan ser puntos de acumulación; es decir, $\operatorname{int}(A) \subset A'$.

También resulta natural que los puntos exteriores de A no sean puntos de acumulación de A. Es decir, que $A' \cap \operatorname{ext}(A) = \varnothing$, o equivalentemente, que $A' \subset \mathbb{R}^n \setminus \operatorname{ext}(A) = \operatorname{int}(A) \cup \operatorname{Fr}(A)$. Estas contenciones establecen una relación muy específica entre los puntos de acumulación de un conjunto A y sus correspondientes puntos interiores y exteriores, lo que no sucede con los puntos frontera. En efecto, el conjunto A del ejemplo 1.9 nos es útil para ilustrar este hecho; en este caso, el $(0,0)$ es un punto de acumulación de

A (ya que es un punto interior de A) y no es un punto frontera de A (por la misma razón); sin embargo, el punto $(2, 2)$ es un punto frontera de A y, como se mostró en ese ejemplo, no es un punto de acumulación de A.

Las contenciones dadas en el párrafo anterior son muy importantes. Por esta razón las dejamos expresadas en la siguiente proposición.

Proposición 1.9 *Si $A \subset \mathbb{R}^n$ es un conjunto arbitrario, entonces*

$$\operatorname{int}(A) \subset A' \subset \operatorname{int}(A) \cup \operatorname{Fr}(A).$$

Demostración. Como es de suponerse, la prueba de esta proposición se deja al lector. ■

Cuando expresamos en términos intuitivos lo que debería de pasar para que un punto $\hat{x} \in \mathbb{R}^n$ fuera un punto de acumulación de un conjunto A, dijimos que cualquier vecindad $B_r(\hat{x})$ (sin importar qué tan grande o pequeño fuera r) tendría que tener "muchos" puntos de A.

Lo siguiente que vamos a mostrar es que, de la definición formal que dimos, en efecto se desprende que esto debe ser así: $\hat{x} \in \mathbb{R}^n$ es un punto de acumulación de A si y sólo si para toda $r > 0$, $B_r(\hat{x}) \cap A$ es un conjunto infinito.

Proposición 1.10 *Sea $A \subset \mathbb{R}^n$. Se satisface que $\hat{x} \in \mathbb{R}^n$ es un punto de acumulación de A si y sólo si para todo $r > 0$ se tiene que $B_r(\hat{x}) \cap A$ es un conjunto infinito.*

Demostración. Para probar que esta propiedad es una condición necesaria del hecho de que $\hat{x}$ es un punto de acumulación de A, procederemos por el método de la contrapuesta.

Supongamos entonces que existe $r > 0$ tal que $B_r(\hat{x}) \cap A$ es un conjunto finito. Si $(B_r(\hat{x}) \setminus \{\hat{x}\}) \cap A = \varnothing$, entonces $\hat{x}$ no es un punto de acumulación de A y por lo tanto ya habremos terminado.

Si $(B_r(\hat{x}) \setminus \{\hat{x}\}) \cap A = \{\hat{x}_1, \ldots, \hat{x}_k\}$, hacemos $r' = \min\{\|\hat{x} - \hat{x}_i\| \mid i \in \{1, \ldots, k\}\}$. Es claro que $r' > 0$ y además, como $r' \leq \|\hat{x} - \hat{x}_i\|$ para toda $i \in$

$\{1, \ldots, k\}$, entonces $(B_{r'}(\hat{x}) \setminus \{\hat{x}\}) \cap A = \varnothing$, con lo que de nuevo concluimos que $\hat{x}$ no es un punto de acumulación de A.

Para probar la suficiencia, basta observar que si $B_r(\hat{x}) \cap A$ es un conjunto infinito para toda $r > 0$, entonces es inmediato que $(B_r(\hat{x}) \setminus \{\hat{x}\}) \cap A \neq \varnothing$, de modo que $\hat{x}$ es un punto de acumulación de A. ■

La proposición anterior tiene un corolario muy interesante, pues establece una condición suficiente para que un conjunto no tenga puntos de acumulación (o si se prefiere, una condición (o consecuencia) necesaria del hecho de que un conjunto sí tenga puntos de acumulación): si un conjunto $A \subset \mathbb{R}^n$ es finito, entonces $A' = \varnothing$ (o equivalentemente: si $A' \neq \varnothing$, entonces A es un conjunto infinito). Escribiremos el corolario de la segunda forma.

Corolario 1.1 *Sea $A \subset \mathbb{R}^n$. Si $A' \neq \varnothing$, entonces A es un conjunto infinito.*

Una pregunta muy importante es si la afirmación recíproca del corolario anterior es cierta: ¿si A es un conjunto infinito, entonces $A' \neq \varnothing$? La respuesta es negativa y el ejemplo es el conjunto A que se da en el problema 25.

La siguiente pregunta es entonces: además de ser infinito, ¿qué otra propiedad debe tener un conjunto A para que se pueda asegurar que al menos tiene un punto de acumulación? La clave para responder esta pregunta nos la dan el conjunto A que acabamos de usar como contraejemplo y el resultado del problema 35. En este problema se establece que, si un conjunto tiene al menos un punto de acumulación (es decir, que $A' \neq \varnothing$), entonces para cualquier cantidad positiva (sin importar cuán pequeña la tomemos), deben existir un par de elementos de A cuya distancia entre ellos sea menor que esa cantidad positiva. Es decir, sin importar que distancia elijamos, siempre debemos poder encontrar elementos en A tales que la distancia entre ellos sea menor que la distancia elegida. Si el lector observa con cuidado, el conjunto A del problema 25 no satisface esta propiedad, y peor aún, en este caso tenemos que $\|\hat{x} - \hat{y}\| \geq 1$ para todo $\hat{x}, \hat{y} \in A$, si $\hat{x} \neq \hat{y}$.

De la discusión anterior se desprende la siguiente pregunta: ¿qué propiedad tiene el conjunto A del problema 25, que además de ser infinito, permite

que la distancia entre cualesquiera dos de sus elementos sea mayor o igual que una cierta cantidad positiva fija? Pues una propiedad del conjunto A que hasta ahora no hemos observado es que, sin importar cuán grande se elija un número $M > 0$, siempre podemos encontrar un $\hat{x} \in A$ tal que $\|\hat{x}\| \geq M$. Más aún, obsérvese que dado un número $M > 0$, sin importar lo "grande" que éste sea, la norma de casi todos los elementos del conjunto A (es decir, todos salvo un número finito de ellos) rebasa a este número M.

Cuando un conjunto A tiene la propiedad de que, sin importar cuán grande se elija un número $M > 0$, siempre podemos encontrar un $\hat{x} \in A$ tal que $\|\hat{x}\| \geq M$, se dice que este conjunto es no acotado. O dicho de manera positiva, cuando sucede lo contrario, decimos que el conjunto está acotado.

Este concepto resultará ser muy importante para la pregunta que nos hicimos con respecto a la proposición recíproca del corolario 1.1, y por esta razón lo dejamos plasmado en la siguiente

Definición 1.11 *Sea $A \subset \mathbb{R}^n$. Decimos que A es un conjunto acotado (o simplemente que A está acotado) si existe $M > 0$ tal que $\|\hat{x}\| \leq M$ para todo $\hat{x} \in A$.*

En términos geométricos, decir que un conjunto A está acotado significa que éste queda contenido en alguna bola (o vecindad) con centro en el origen (observe que, de la definición anterior, se desprende que $A \subset B_{M+1}(\hat{0})$).

Como seguramente el lector ya habrá sospechado, para que podamos asegurar que un conjunto infinito A tiene al menos un punto de acumulación, será suficiente que éste también sea acotado. Este es un resultado tan importante, que hasta tiene nombre: teorema de Bolzano-Weierstrass[5]. La

[5]El teorema de Bolzano-Weierstrass lleva el nombre de los matemáticos *Bernard Placidus Johann Gonzal Nepomuk Bolzano (Praga, Bohemia (actual República Checa), 5 de octubre de 1781 - ídem, 18 de diciembre de 1848)*, y *Karl Theodor Wilhelm Weierstraß (escrito Weierstrass cuando no está disponible el caracter "ß") (Ostenfelde, 31 de octubre de 1815 -Berlín, 19 de febrero de 1897)*. El primero, conocido como Bernard Bolzano, fue un matemático, lógico, filósofo y teólogo bohemio que escribió en alemán y que realizó importantes contribuciones a las matemáticas y a la Teoría del conocimiento. El segundo, conocido como Karl Weierstrass, fue un matemático alemán al que se suele citar como el "padre del análisis moderno". En realidad, este teorema fue demostrado por primera vez

prueba de este teorema requiere que contemos con la versión generalizada (a $\mathbb{R}^n$) de un conocido teorema de los números reales: el teorema de los intervalos anidados. Por esta razón, primero definiremos todo lo necesario para formular dicha generalización, la probaremos, y finalmente la usaremos para demostrar el teorema de Bolzano-Weierstrass.

Lo primero que haremos será generalizar el concepto de intervalo cerrado, de la siguiente manera.

Definición 1.12 *Dados $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ tales que $a_i \leq b_i$ para $i = 1, \ldots, n$, decimos que el conjunto*

$$R = [a_1, b_1] \times \cdots \times [a_n, b_n] = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i, i \in \{1, \ldots, n\}\}$$

es un rectángulo cerrado. Asimismo, definimos la diagonal de R, que denotamos por $\operatorname{diag}(R)$*, como*

$$\operatorname{diag}(R) := \|(b_1, \ldots, b_n) - (a_1, \ldots, a_n)\|.$$

Establecemos las propiedades más elementales de este tipo de conjuntos en el siguiente

Lema 1.1 *Sea $R = [a_1, b_1] \times \cdots \times [a_n, b_n] = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i, i = 1, \ldots, n\}$ un rectángulo cerrado. Se satisfacen las siguientes afirmaciones:*

1. *R es un conjunto cerrado,*
2. *si $\hat{x}, \hat{y} \in R$, entonces $\|\hat{x} - \hat{y}\| \leq \operatorname{diag}(R)$,*
3. *si $\hat{x} \in R$ y $\operatorname{diag}(R) < r$, entonces $R \subset B_r(\hat{x})$,*
4. *si $R' = [a'_1, b'_1] \times \cdots \times [a'_n, b'_n]$ se tiene que: $R' \subset R$ si y sólo si $[a'_i, b'_i] \subset [a_i, b_i]$ para toda $i \in \{1, \ldots, n\}$.*

en 1817 por Bolzano, como un lema en la demostración del teorema de valor intermedio. Unos cincuenta años más tarde, el resultado fue identificado como significativo por derecho propio, y demostrado una vez más por Weierstrass. Desde entonces se ha convertido en un teorema fundamental del análisis. (fuente: Wikipedia).

El lector estará de acuerdo en que todas estas propiedades son muy sencillas de probar, razón por la cual se deja que él las haga.

Habiendo dado la definición y lema anteriores, estamos en condiciones de formular y probar lo que llamaremos el *teorema de los rectángulos anidados*, de la siguiente manera.

Teorema 1.2 (de los rectángulos anidados) *Si $\{R_k\}$ es una sucesión anidada de rectángulos cerrados (es decir, $R_{k+1} \subset R_k$ para toda $k \in \mathbb{N}$), entonces*

$$\bigcap_{k=1}^{\infty} R_k \neq \varnothing.$$

Si además se tiene que $\lim_{k\to\infty} \operatorname{diag}(R_k) = 0$, entonces

$$\bigcap_{k=1}^{\infty} R_k = \{\hat{x}_0\}.$$

para algún $\hat{x}_0 \in \mathbb{R}^n$.

Demostración. Esta prueba estará basada en el correspodiente teorema de los intervalos anidados. Supongamos que para cada $k \in \mathbb{N}$ se tiene que

$$R_k = \left[a_1^{(k)}, b_1^{(k)}\right] \times \cdots \times \left[a_n^{(k)}, b_n^{(k)}\right].$$

Por el inciso 4 del lema 1.1 sabemos que, para cada $i \in \{1, \ldots, n\}$, la sucesión

$$\left\{I_k^{(i)} = \left[a_i^{(k)}, b_i^{(k)}\right]\right\}$$

es una sucesión de intervalos (cerrados) anidados, de tal forma que $\bigcap_{k=1}^{\infty} I_k^{(i)} \neq \varnothing$. Por tanto, si para cada $i \in \{1, \ldots, n\}$ elegimos $x_i \in \bigcap_{k=1}^{\infty} I_k^{(i)}$, entonces se verifica que $\hat{x} = (x_1, \ldots, x_n) \in \bigcap_{k=1}^{\infty} R_k$. En efecto, para cada $i \in \{1, \ldots, n\}$ se tiene que

$$x_i \in I_k^{(i)} = \left[a_i^{(k)}, b_i^{(k)}\right],$$

de tal forma que

$$\hat{x} = (x_1, \ldots, x_n) \in \left[a_1^{(k)}, b_1^{(k)}\right] \times \cdots \times \left[a_n^{(k)}, b_n^{(k)}\right] = R_k.$$

Como lo anterior es válido para cada $k \in \mathbb{N}$, entonces $\hat{x} \in \bigcap_{k=1}^{\infty} R_k$, es decir que $\bigcap_{k=1}^{\infty} R_k \neq \varnothing$.

Supongamos ahora que $\lim_{k\to\infty} \operatorname{diag}(R_k) = 0$ y que $\hat{x}, \hat{y} \in \bigcap_{k=1}^{\infty} R_k$. Por el inciso 2 del lema 1.1 sabemos que

$$0 \leq \|\hat{x} - \hat{y}\| \leq \operatorname{diag}(R_k),$$

de tal forma que, si en las desigualdades anteriores tomamos el límite cuando $k \to \infty$, obtenemos que

$$0 \leq \|\hat{x} - \hat{y}\| \leq 0,$$

es decir, que $\|\hat{x} - \hat{y}\| = 0$ y por tanto que $\hat{x} = \hat{y}$. Esto prueba que todos los elementos de $\bigcap_{k=1}^{\infty} R_k$ son iguales, lo que significa que dicho conjunto consta de un sólo punto. ■

Ya casi todo está listo para poder formular y probar el teorema de Bolzano-Weierstrass; sólo resta mencionar la forma en que, en dicha demostración, subdividiremos a un rectángulo $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$. Simplemente partimos a cada intervalo coordenado $[a_i, b_i]$ en los subintervalos $[a_i, (a_i + b_i)/2]$ y $[(a_i + b_i)/2, b_i]$ y construimos todos los rectángulos que tengan como i-ésimo intervalo coordenado a alguno de estos subintervalos, para $i \in \{1, \ldots, n\}$. Con este procedimiento obtenemos 2^n rectángulos cerrados R_k cuyas propiedades, que por ahora nos interesa destacar, son las siguientes: $R_k \subset R$, $R = \bigcup_{k=1}^{2^n} R_k$ y $\operatorname{diag}(R_k) = \operatorname{diag}(R)/2$ (propiedades que el lector podrá probar muy fácilmente).

Una vez hecho lo anterior, formularemos el teorema de Bolzano-Weierstrass, no sin antes mencionar que el procedimiento que usaremos en su demostración hasta tiene un nombre: *la cacería del león.*

Teorema 1.3 (de Bolzano-Weierstrass) *Si $A \subset \mathbb{R}^n$ es un conjunto infinito y acotado, entonces A tiene al menos un punto de acumulación (es decir, $A' \neq \varnothing$).*

Demostración. Sea $M > 0$ tal que $\|\hat{x}\| \leq M$ para toda $\hat{x} \in A$. Si hacemos $R = [-M, M] \times \cdots \times [-M, M]$, tenemos entonces que $A \subset R$. Ahora subdividimos a R en la forma que mencionamos antes.

Dado que $A \subset R$, que R es la unión de todos los rectángulos inducidos por la subdivisión que hicimos, y que A es infinito, se debe tener entonces que en alguno de estos rectángulos inducidos, al que llamaremos R_1, debe haber una infinidad de puntos de A. Es decir, $R_1 \subset R$ es un rectángulo cerrado tal que $A \cap R_1$ es un conjunto infinito y

$$\operatorname{diag}(R_1) = \operatorname{diag}(R)/2 = (2\sqrt{n}M)/2 = \sqrt{n}M.$$

Ahora subdividimos a R_1 en la forma que mencionamos antes. Dado que $A \cap R_1 \subset R_1$, que R_1 es la unión de todos los rectángulos inducidos por la subdivisión que le hicimos, y que $A \cap R_1$ es infinito, entonces en alguno de estos rectángulos inducidos, al que llamaremos R_2, debe haber una infinidad de puntos de $A \cap R_1$. Es decir, $R_2 \subset R_1$ es un rectángulo cerrado tal que

$$(A \cap R_1) \cap R_2 = A \cap (R_1 \cap R_2) = A \cap R_2$$

es un conjunto infinito y

$$\operatorname{diag}(R_2) = \operatorname{diag}(R_1)/2 = \operatorname{diag}(R)/2^2 = (2\sqrt{n}M)/2^2 = \sqrt{n}M/2.$$

Como el lector ya habrá notado, este es un procedimiento que podemos seguir indefinidamente, es decir, es un proceso inductivo. En efecto, si para una cierta $k \in \mathbb{N}$ ya hemos construido un rectángulo cerrado R_k con las siguientes propiedades:

1. $R_k \subset R_{k-1} \subset \cdots \subset R_1 \subset R$,
2. $A \cap R_k$ es un conjunto infinito, y
3. $\operatorname{diag}(R_k) = \frac{\sqrt{n}M}{2^{k-1}}$,

entonces podemos construir un rectángulo cerrado R_{k+1}, que tenga las propiedades equivalentes. ¿Cómo? De la siguiente manera: subdividimos a R_k en la forma en que hemos venido haciéndolo. Dado que $A \cap R_k \subset R_k$, que

R_k es la unión de todos los rectángulos inducidos por la subdivisión que le hicimos, y que $A \cap R_k$ es infinito, en alguno de los rectángulos inducidos por esta subdivisión, al que llamaremos R_{k+1}, debe haber una infinidad de puntos de $A \cap R_k$. Es decir, R_{k+1} será un rectángulo cerrado que satisface las propiedades deseadas:

1. $R_{k+1} \subset R_k$,
2. $(A \cap R_k) \cap R_{k+1} = A \cap (R_k \cap R_{k+1}) = A \cap R_{k+1}$ es un conjunto infinito, y
3. $\operatorname{diag}(R_{k+1}) = \operatorname{diag}(R_k)/2 = (\sqrt{n}M/2^{k-1})/2 = \sqrt{n}M/2^k$

(ver figura 1.14).

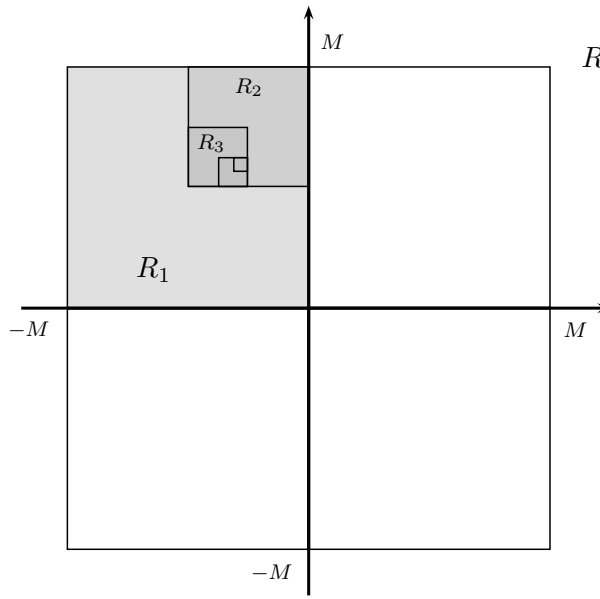


Figura 1.14: La *cacería del león*.

Con base en este procedimiento obtenemos una sucesión (infinita) de rectángulos cerrados $\{R_k\}$ que satisfacen las propiedades que enlistamos

anteriormente. De esta forma, por el teorema 1.2 (de los rectángulos anidados), sabemos que

$$\bigcap_{k=1}^{\infty} R_k = \{\hat{x}_0\}$$

para algún $\hat{x}_0 \in \mathbb{R}^n$.

Aseguramos que $\hat{x}_0$ es un punto de acumulación de A. Sea $r > 0$; como

$$\lim_{k\to\infty} \operatorname{diag}(R_k) = 0,$$

existe $N \in \mathbb{N}$ tal que si $k \geq N$, entonces $0 \leq \operatorname{diag}(R_k) < r$. Por otra parte, como $\hat{x}_0 \in R_k$ para toda $k \in \mathbb{N}$, entonces por el inciso 3 del lema 1.1 se tiene que $R_k \subset B_r(\hat{x}_0)$, de tal forma que $A \cap R_k \subset A \cap B_r(\hat{x}_0)$, y como $A \cap R_k$ es un conjunto infinito, podemos concluir que $(B_r(\hat{x}_0) \setminus \{\hat{x}_0\}) \cap A \neq \varnothing$ y por lo tanto que $\hat{x}_0$ es un punto de acumulación de A. ■

1.5.4. Conjuntos conexos

Concluiremos esta sección mostrando de qué forma los conceptos que hasta ahora hemos desarrollado, se pueden usar para definir de manera muy precisa un hecho geométrico que es intuitivamente muy claro: ¿cuándo un conjunto ¿está formado de una sola pieza"? Es decir, ¿cuándo un conjunto no está "roto"?

En realidad lo que haremos será establecer cuándo se puede decir que un conjunto está "roto" o "separado" (y, consecuentemente, los conjuntos que son de una sola pieza (o conexos, como suele llamárseles) serán los que no están "rotos" o "separados").

Intuitivamente, un conjunto $A \subset \mathbb{R}^n$ estará "roto" si se puede poner como la unión de otros dos conjuntos B y C, para los cuales hay alguna forma de decir que están "separados".

Tal vez la primera idea que se nos venga a la cabeza para decir que dos conjuntos están separados es que simplemente sean ajenos (y claro, ¡diferentes del vacío!). Desafortunadamente, aún cuando dos conjuntos sean

ajenos, estos pueden embonar (o empatar) muy bien, de tal forma que al unirlos formen un conjunto que no esté "roto".

Tal es el caso de los siguientes conjuntos: $B = [0, 1/2] \times [0, 1]$ y $C = (1/2, 1] \times [0, 1]$ (ver figura 1.15) cuya unión nos da el conjunto $A = [0, 1] \times [0, 1]$, el cual no es el tipo de conjunto del que se pudiera decir que está "roto". De hecho, cualquier conjunto A con más de un punto (esté o no esté "roto") podemos expresarlo como la unión de dos conjuntos ajenos; bastaría con tomar un subconjunto propio B (y no vacío) de A ($\varnothing \neq B \subsetneq A$) y $C = A \setminus B$ para que el conjunto A quedara expresado de esta forma.

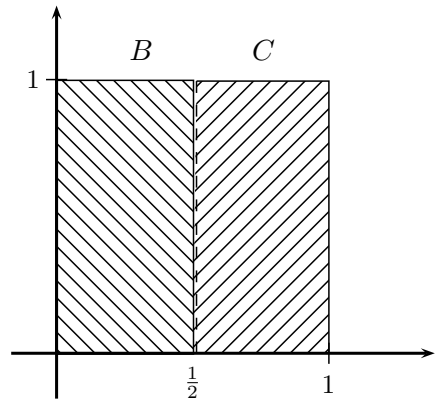


Figura 1.15: El conjunto $A = [0, 1] \times [0, 1]$, que no está "roto", es la unión de los conjuntos ajenos $B = [0, 1/2] \times [0, 1]$ y $C = (1/2, 1] \times [0, 1]$.

Por el ejemplo (y la observación) anterior concluimos que pedir que dos conjuntos sólo sean ajenos, no es una buena forma de decir que éstos están separados. Pero no hay que desanimarse, pues la clave de nuestro problema se encuentra en la segunda clasificación que hicimos de los puntos de $\mathbb{R}^n$ con respecto a un conjunto A. En efecto, como se recordará, intuitivamente los puntos de acumulación de un conjunto se pueden interpretar como los puntos que están "pegados" a A, de tal forma que si un conjunto no sólo es ajeno a A sino que tampoco tiene puntos que están "pegados" a A (es decir, es ajeno a A'), entonces sí que podremos decir que dicho conjunto está separado de

A. De esta forma, si B y son dos conjuntos tales que $B \cap C = \varnothing$) y además $B \cap C' = \varnothing$, entonces podemos decir que B está separado de C. Si también sucede que $B' \cap C = \varnothing$, entonces podemos decir que C está separado de B, o simplemente, que B y C están separados.

Con base en la discusión anterior definiremos lo que significa que dos conjuntos estén separados. Como seguramente el lector ya sospecha, en esta definición aparecerá la unión de un conjunto y su correspondiente conjunto de puntos de acumulación, es decir $A \cup A'$. Por esta razón, adelantaremos que el lector probará en el inciso (d) del ejercicio 27 que esta unión es igual a la cerradura de A, es decir que $A \cup A' = \bar{A}$, y que es como nos referiremos a ella en la siguiente

Definición 1.13 *Sean $B, C \subset \mathbb{R}^n$. Decimos que B y C están separados si $B \cap \bar{C} = \varnothing = \bar{B} \cap C$.*

Si bien es cierto que en la discusión previa a esta definición hicimos notar que el hecho de que dos conjuntos sean ajenos no es suficiente para que estén separados, si los conjuntos son de cierto tipo, entonces el que sean ajenos sí es suficiente para que estén separados. Estos casos los dejaremos expresados en la siguiente proposición, cuya prueba quedará a cargo del lector.

Proposición 1.11 *Sean $B, C \subset \mathbb{R}^n$.*

1. *Si B y C son abiertos, entonces B y C están separados si y sólo si B y C son ajenos (es decir, $B \cap C = \varnothing$).*

2. *Si B y C son cerrados, entonces B y C están separados si y sólo si B y C son ajenos (es decir, $B \cap C = \varnothing$).*

Como habíamos adelantado, una vez que contamos con este concepto es muy sencillo definir lo que significa que un conjunto esté "roto", o que sea *disconexo*, que es el término que realmente se usa.

Definición 1.14 *Sea $A \subset \mathbb{R}^n$. Decimos que A es un conjunto disconexo si existen $B, C \subset \mathbb{R}^n$, no vacíos, tales que:*

1. *$A = B \cup C$, y*

2. B y C están separados, es decir $B \cap \bar{C} = \varnothing = \bar{B} \cap C$.

Como también ya habíamos mencionado, un conjunto será *conexo* si no es disconexo (o dicho con toda propiedad, si es no disconexo), lo que formalizamos en la (casi) última definición de este capítulo.

Definición 1.15 *Sea $A \subset \mathbb{R}^n$. Decimos que A es un conjunto conexo si A es no disconexo.*

La definición anterior amerita un comentario importante: *el concepto de conexidad se define con base en una negación.* Esta es una característica que es importante resaltar, puesto que este tipo de conceptos son un poco más difíciles de manejar. Específicamente, si se quisiera demostrar *directamente* que un conjunto A es conexo, habría que *demostrar que no existen* $B, C \subset \mathbb{R}^n$, no vacíos y separados, tales que $A = B \cup C$. Como seguramente el lector a estas alturas ya habrá aprendido, demostrar que algo no existe siempre suele ser más difícil; casi siempre lo más práctico es suponer que ese algo sí existe, para después tratar de llegar a una contradicción. Es decir, proceder por contradicción.

Por esta misma razón, encontrar condiciones suficientes y/o necesarias para que un conjunto sea conexo adquiere particular importancia, y eso es justo lo que vamos a hacer en lo que resta de esta sección.

Para iniciar esta tarea, empezaremos por establecer una condición necesaria y suficiente para que un conjunto abierto sea conexo, para lo cual nos será útil dar la siguiente

Definición 1.16 *Sean $\hat{x}, \hat{y} \in \mathbb{R}^n$. Definimos el segmento (de recta) que une a $\hat{x}$ con $\hat{y}$, y que denotamos por $[\hat{x}, \hat{y}]$, como el conjunto dado por*

$$[\hat{x}, \hat{y}] := \{\hat{x} + t(\hat{y} - \hat{x}) = (1-t)\hat{x} + t\hat{y} \in \mathbb{R}^n \mid 0 \leq t \leq 1\}.$$

El primer resultado que probaremos, y que nos será muy útil en todo lo relacionado con la conexidad es "geométricamente" muy claro: si dos conjuntos $B, C \subset \mathbb{R}^n$ están separados y tomamos un punto $\hat{x} \in B$ y otro punto $\hat{y} \in C$, es de esperarse que el segmento que une a estos puntos no esté contenido en la unión de B y C. Es decir que $[\hat{x}, \hat{y}] \nsubseteq B \cup C$. Este hecho lo probaremos en el siguiente

Lema 1.2 *Sean $B, C \subset \mathbb{R}^n$ tales que $\hat{x} \in B$ y $\hat{y} \in C$. Si B y C están separados, entonces el segmento que une a $\hat{x}$ y $\hat{y}$ no está contenido en la unión de B y C, es decir $[\hat{x}, \hat{y}] \nsubseteq B \cup C$.*

Demostración. Construimos el siguiente conjunto:

$$S = \{s \in [0,1] \mid \hat{x} + t(\hat{y} - \hat{x}) \in (\bar{C})^c \text{ para toda } t \in [0, s]\}.$$

Observe que este conjunto es no vacío, pues tomando $s = 0$ en la definición del conjunto S obtenemos el punto $\hat{x}$ que pertenece a $B \subset (\bar{C})^c$, en donde esta última contención se cumple ya que $B \cap \bar{C} = \varnothing$, lo cual a su vez es cierto puesto que B y C están separados. Por otra parte, S está contenido en el intervalo $[0, 1]$ y por lo tanto está acotado superiormente por el 1 (e inferiormente por el 0, lo que por ahora no resulta muy importante). De esta forma, por la propiedad del supremo de los números reales, existe $\alpha \in \mathbb{R}$ tal que $\alpha = \sup(S)$. Dado que $S \subset [0, 1]$, se tiene que $\alpha \in [0, 1]$, de modo que si $\hat{x}_0 = \hat{x} + \alpha(\hat{y} - \hat{x})$, entonces $\hat{x}_0 \in [\hat{x}, \hat{y}]$. Con respecto a este número α, es importante hacer notar que se satisface la siguiente contención:

$$[0, \alpha) \subset S. \tag{1.11}$$

En efecto, si $0 \leq s < \alpha$, de la definición de supremo sabemos que existe $s' \in S$ tal que $s < s' < \alpha$, de modo que, por la forma en que está definido el conjunto S, se tiene que $\hat{x} + t(\hat{y} - \hat{x}) \in (\bar{C})^c$ para toda $t \in [0, s']$ y por lo tanto para toda $t \in [0, s] \subset [0, s']$, por lo que se concluye que $s \in S$.

Otro hecho que podemos asegurar es que $\hat{x}_0 \notin B$. En efecto, si $\hat{x}_0$ estuviera en B, dado que $B \subset (\bar{C})^c$ (por la razón que ya explicamos), entonces $\alpha < 1$ y $\hat{x}_0 \in (\bar{C})^c$. Ahora, como $\bar{C}$ es un conjunto cerrado (inciso 2 de la proposición 1.8), entonces $(\bar{C})^c$ es un conjunto abierto de modo que existe $r > 0$ tal que $B_r(\hat{x}_0) \subset (\bar{C})^c$. Nótese ahora que, si

$$r' = \text{mín}\left\{\frac{1-\alpha}{2}, \frac{r}{2\|\hat{x} - \hat{y}\|}\right\},$$

entonces para toda $\alpha \leq t \leq \alpha + r'$ se tiene que

$$\begin{aligned}\|\hat{x}_0 - (\hat{x} + t(\hat{y} - \hat{x}))\| &= \|(\hat{x} + \alpha(\hat{y} - \hat{x})) - (\hat{x} + t(\hat{y} - \hat{x}))\| \\ &= \|(\alpha - t)(\hat{y} - \hat{x}))\| \\ &= (t - \alpha)\|\hat{x} - \hat{y}\| \\ &\leq r'\|\hat{x} - \hat{y}\| \\ &\leq \frac{r}{2} \\ &< r\end{aligned}$$

y por lo tanto $\hat{x} + t(\hat{y} - \hat{x}) \in B_r(\hat{x}_0) \subset (\bar{C})^c$. De esta contención, y de la contención 1.11 concluimos que $\hat{x}+t(\hat{y}-\hat{x}) \in (\bar{C})^c$ para toda $t \in [0, \alpha+r'] \subset [0, 1]$ y por lo tanto que $\alpha + r' \in S$, pertenencia que contradice el hecho de que $\alpha = \sup(S)$.

Si $\hat{x}_0 \notin C$ ya habremos terminado, pues entonces $\hat{x}_0 \notin B \cup C$. Pero sí puede suceder que $\hat{x}_0 \in C$ (¿podría dar el lector un ejemplo en el que esto fuera cierto?). En este caso mostraremos que podemos obtener otro punto $\hat{x}'_0 \in [\hat{x}, \hat{y}]$ tal que $\hat{x}'_0 \notin B \cup C$, de la siguiente forma: si $\hat{x}_0 \in C$, dado que $\bar{B} \cap C = \varnothing$ (porque B y C están separados), entonces $0 < \alpha$ y $\hat{x}_0 \in (\bar{B})^c$, y como este es un conjunto abierto, existe $r > 0$ tal que $B_r(\hat{x}_0) \subset (\bar{B})^c$. Sea

$$t' = \text{mín}\left\{\frac{\alpha}{2}, \frac{r}{2\|\hat{x} - \hat{y}\|}\right\}.$$

Entonces, como $0 < \alpha - t' < \alpha \leq 1$, si tomamos $\hat{x}'_0 = \hat{x} + (\alpha - t')(\hat{y} - \hat{x}) \in [\hat{x}, \hat{y}]$ tenemos, por la contención 1.11, que $\alpha - t' \in S$ y por lo tanto que $\hat{x}'_0 \in (\bar{C})^c$, de donde concluimos que $\hat{x}'_0 \notin C$. Por otra parte, como

$$\begin{aligned}\|\hat{x}_0 - \hat{x}'_0\| &= \big\|(\hat{x} + \alpha(\hat{y} - \hat{x})) - ((\hat{x} + (\alpha - t')(\hat{y} - \hat{x})))\big\| \\ &= \big\|t'(\hat{y} - \hat{x})\big\| \\ &= t'\|\hat{x} - \hat{y}\| \\ &\leq \frac{r}{2} \\ &< r,\end{aligned}$$

se tiene que $\hat{x}_0' \in B_r(\hat{x}_0) \subset (\bar{B})^c$ de donde ahora concluimos que $\hat{x}_0' \notin B$. Es decir, $\hat{x}_0' \notin B \cup C$. ■

Con base en este lema vamos a obtener resultados muy importantes; con el primero de ellos lograremos probar que una clase muy grande y muy importante de conjuntos son conexos. Nos referimos a los llamados *conjuntos convexos*, los cuales definimos a continuación.

Definición 1.17 *Sea $A \subset \mathbb{R}^n$. Decimos que A es un conjunto convexo si para cada par de puntos $\hat{x}, \hat{y} \in A$ se tiene que $[\hat{x}, \hat{y}] \subset A$.*

Antes de probar el resultado que anunciamos, mostraremos algunos ejemplos de conjuntos convexos.

Ejemplo 1.10 *Los siguientes conjuntos son ejemplos de conjuntos convexos:*

1. *El espacio total $\mathbb{R}^n$. Esto es un hecho inmediato, puesto que estamos hablando del conjunto que constituye el "universo" sobre el cual se está trabajando. De esta forma, si $\hat{x}, \hat{y} \in \mathbb{R}^n$, por supuesto que $[\hat{x}, \hat{y}] \subset \mathbb{R}^n$.*

2. *Para todo $\hat{x}, \hat{y} \in \mathbb{R}^n$, el segmento que los une: $[\hat{x}, \hat{y}]$. Observe que, si tomamos $\hat{x} + s(\hat{y} - \hat{x})$ y $\hat{x} + s'(\hat{y} - \hat{x})$, con $s, s' \in [0, 1]$, es decir $\hat{x} + s(\hat{y} - \hat{x})$, $\hat{x} + s'(\hat{y} - \hat{x}) \in [\hat{x}, \hat{y}]$, entonces se tiene que*

$$\begin{aligned}(\hat{x} + s(\hat{y} - \hat{x})) + t(\hat{x} + s'(\hat{y} - \hat{x}) - (\hat{x} + s(\hat{y} - \hat{x})))& \\ = \hat{x} + (s + t(s' - s))(\hat{y} - \hat{x})& \\ = \hat{x} + ((1 - t)s + ts')(\hat{y} - \hat{x}).&\end{aligned}$$

   Como $(1 - t)s + ts' \in [0, 1]$ para toda $t \in [0, 1]$, concluimos que $[\hat{x} + s(\hat{y} - \hat{x}), \hat{x} + s'(\hat{y} - \hat{x})] \subset [\hat{x}, \hat{y}]$. Es decir que $[\hat{x}, \hat{y}]$ es convexo.

3. *La bola (o vecindad) de radio $r > 0$ con centro en el punto $\hat{x}_0$ ($B_r(\hat{x}_0)$), en donde $r > 0$ es cualquier cantidad positiva y $\hat{x}_0$ es un punto arbitrario en $\mathbb{R}^n$. En efecto, si $\hat{x}, \hat{y} \in B_r(\hat{x}_0)$ y $t \in [0, 1]$ se tiene que*

$$\begin{aligned}\|\hat{x}_0 - (\hat{x} + t(\hat{y} - \hat{x}))\| &= \|(\hat{x}_0 + t(\hat{x}_0 - \hat{x}_0)) - (\hat{x} + t(\hat{y} - \hat{x}))\| \\ &= \|(t\hat{x}_0 + (1-t)\hat{x}_0) - (t\hat{y} + (1-t)\hat{x})\| \\ &= \|t(\hat{x}_0 - \hat{y}) + (1-t)(\hat{x}_0 - \hat{x})\| \\ &\leq t\,\|\hat{x}_0 - \hat{y}\| + (1-t)\,\|\hat{x}_0 - \hat{x}\| \\ &< tr + (1-t)r \\ &= r,\end{aligned}$$

lo que prueba que $[\hat{x}, \hat{y}] \subset B_r(\hat{x}_0)$.

4. *Cualquier rectángulo cerrado* $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$. *Sean* $\hat{x} = (x_1, \ldots, x_n), \hat{y} = (y_1, \ldots, y_n) \in R$ *y sea* $t \in [0,1]$. *Como* $a_i \leq x_i, y_i \leq b_i$ *y* $t \in [0,1]$, *entonces* $ta_i \leq ty_i \leq tb_i$ *y* $(1-t)a_i \leq (1-t)x_i \leq (1-t)b_i$, *de tal forma que*

$$a_i = ta_i + (1-t)a_i \leq ty_i + (1-t)x_i \leq tb_i + (1-t)b_i = b_i.$$

Es decir, tenemos que

$$a_i \leq ty_i + (1-t)x_i = x_i + t(y_i - x_i) \leq b_i$$

para cada $i \in \{1, \ldots, n\}$, *y por lo tanto que* $\hat{x} + t(\hat{y} - \hat{x}) = (x_1 + t(y_1 - x_1), \ldots, x_n + t(y_n - x_n)) \in R$ *para toda* $t \in [0,1]$, *lo que significa que* $[\hat{x}, \hat{y}] \subset R$.

El resultado que probaremos a continuación tendrá como corolario un hecho muy importante: todos los conjuntos mencionados en el ejemplo anterior serán conjuntos conexos.

Proposición 1.12 *Sea* $A \subset \mathbb{R}^n$. *Si* A *es convexo, entonces* A *es conexo.*

Demostración. Supongamos que A no es conexo. Entonces existen $B, C \subset \mathbb{R}^n$ no vacíos y separados, tales que $A = B \cup C$. Dado que B y C son no vacíos, elegimos $\hat{x} \in B$ y $\hat{y} \in C$. Por el lema 1.2 tenemos entonces que

$[\hat{x}, \hat{y}] \nsubseteq B \cup C = A$ lo cual contradice el hecho de que A es convexo. Por lo tanto, A es conexo. ■

Una pregunta que siempre nos tenemos que hacer es si el recíproco de una proposición es cierto. En el caso de la proposición anterior, la respuesta es negativa. Para poder dar un ejemplo de este hecho, será necesario ampliar la familia de conjuntos para los cuales hayamos probado que son conexos. Justo eso es lo que haremos en la siguiente proposición que probaremos, la cual nos proporciona una condición necesaria y suficiente para que un conjunto abierto sea conexo.

Para facilitar la redacción de su prueba, introducimos el siguiente concepto: dados $\hat{x}, \hat{y}, \hat{x}_1, \ldots, \hat{x}_k \in \mathbb{R}^n$, diremos que el conjunto $[\hat{x}, \hat{x}_1] \cup [\hat{x}_1, \hat{x}_2] \cup \cdots \cup [\hat{x}_k, \hat{y}]$ es *una poligonal* que une a $\hat{x}$ con $\hat{y}$ (o que empieza en $\hat{x}$ y termina en $\hat{y}$, o que tiene extremos $\hat{x}$ y $\hat{y}$).

Proposición 1.13 *Sea $A \subset \mathbb{R}^n$ abierto y no vacío. A es conexo si y sólo si para cada par de puntos $\hat{x}, \hat{y} \in A$, existen $\hat{x}_1, \ldots, \hat{x}_k \in A$ tales que $[\hat{x}, \hat{x}_1] \cup [\hat{x}_1, \hat{x}_2] \cup \cdots \cup [\hat{x}_k, \hat{y}] \subset A$. Es decir, A es conexo si y sólo si para cada par de puntos $\hat{x}, \hat{y} \in A$ existe una poligonal contenida en A que une a $\hat{x}$ con $\hat{y}$.*

Demostración. Supongamos que A es conexo y sea $\hat{x}_0 \in A$ un punto fijo. Nótese que nuestra demostración se reduce a demostrar que para cualquier $\hat{x} \in A$, siempre existe una poligonal que une a $\hat{x}_0$ con $\hat{x}$ (¿por qué?). Para lograr esto, definimos los siguientes conjuntos:

$$U = \{\hat{x} \in A \mid \textit{existe} \text{ una poligonal contenida en } A \text{ que une a } \hat{x}_0 \text{ con } \hat{x}\}$$

y

$$V = \{\hat{x} \in A \mid \textit{no existe} \text{ una poligonal contenida en } A \text{ que une a } \hat{x}_0 \text{ con } \hat{x}\}.$$

Observe que nuestro objetivo es probar que $U = A$.

Como se prueba fácilmente, los conjuntos U y V tienen las siguientes propiedades:

1. $A = U \cup V$,

2. $U \cap V = \varnothing$, y

3. $U \neq \varnothing$ (pues $\hat{x}_0 \in U$).

Lo que ahora vamos a demostrar es que U y V son conjuntos abiertos. En efecto, si $\hat{x} \in U \subset A$, como A es un conjunto abierto, sabemos que existe $r > 0$ tal que $B_r(\hat{x}) \subset A$; aseguramos que $B_r(\hat{x}) \subset U$ ya que si $\hat{y} \in B_r(\hat{x})$ sabemos que $[\hat{x}, \hat{y}] \subset B_r(\hat{x}) \subset A$ (las bolas son convexas) de tal forma que si $[\hat{x}_0, \hat{x}_1] \cup [\hat{x}_1, \hat{x}_2] \cup \cdots \cup [\hat{x}_k, \hat{x}]$ es una poligonal contenida en A que une a $\hat{x}_0$ con $\hat{x}$, entonces $[\hat{x}_0, \hat{x}_1] \cup [\hat{x}_1, \hat{x}_2] \cup \cdots \cup [\hat{x}_k, \hat{x}] \cup [\hat{x}, \hat{y}]$ es una poligonal contenida en A que une a $\hat{x}_0$ con $\hat{y}$. De lo anterior concluimos que $\hat{y} \in U$, es decir, $B_r(\hat{x}) \subset U$ y por lo tanto que U es un conjunto abierto.

Análogamente, si $\hat{x} \in V \subset A$, como A es un conjunto abierto, sabemos que existe $r > 0$ tal que $B_r(\hat{x}) \subset A$; aseguramos que $B_r(\hat{x}) \subset V$. Si esta contención no se cumpliera, existiría $\hat{y} \in B_r(\hat{x}) \subset A$ tal que $\hat{y} \notin V$ y por tanto se tendría que $\hat{y} \in U$ en cuyo caso existiría $[\hat{x}_0, \hat{x}_1] \cup [\hat{x}_1, \hat{x}_2] \cup \cdots \cup [\hat{x}_k, \hat{y}]$, una poligonal contenida en A que une a $\hat{x}_0$ con $\hat{y}$. Como $[\hat{y}, \hat{x}] \subset B_r(\hat{x}) \subset A$, entonces $[\hat{x}_0, \hat{x}_1] \cup [\hat{x}_1, \hat{x}_2] \cup \cdots \cup [\hat{x}_k, \hat{y}] \cup [\hat{y}, \hat{x}]$ sería una poligonal contenida en A que uniría a $\hat{x}_0$ con $\hat{x}$, contradiciendo el hecho de que $\hat{x} \in V$. Así pues, se debe tener que $B_r(\hat{x}) \subset V$, con lo cual concluimos que V también es abierto.

Resumiendo, tenemos que A es la unión de dos conjuntos abiertos y ajenos U y V, que por la proposición 1.11 están separados. Por estas razones, si además también se cumpliera que $V \neq \varnothing$, obtendríamos que A no es conexo, lo que sería una contradicción. De esta forma se debe tener que $V = \varnothing$ y por tanto que $A = U$, que es justo lo que se quería probar.

La condición de suficiencia para la conexidad es una consecuencia inmediata del lema 1.2. Si A no fuera conexo sabríamos que existen B y C conjuntos no vacíos y separados tales que $A = B \cup C$. De esta forma, si elegimos $\hat{x} \in B$, $\hat{y} \in C$ y tomamos $[\hat{x}, \hat{x}_1] \cup [\hat{x}_1, \hat{x}_2] \cup \cdots \cup [\hat{x}_k, \hat{y}]$ una poligonal que une a estos puntos, entonces aseguramos que $[\hat{x}, \hat{x}_1] \cup [\hat{x}_1, \hat{x}_2] \cup \cdots \cup [\hat{x}_k, \hat{y}] \nsubseteq A$. En efecto, nótese que si $\hat{x}_i \notin A$ para alguna $i \in \{1, \ldots, k\}$, entonces $[\hat{x}, \hat{x}_1] \cup [\hat{x}_1, \hat{x}_2] \cup \cdots \cup [\hat{x}_k, \hat{y}] \nsubseteq A$. Si $\hat{x}_i \in A$ para toda $i \in \{1, \ldots, k\}$, llamando $\hat{x}_0 = \hat{x}$ y $\hat{x}_{k+1} = \hat{y}$ podemos asegurar que existe $i \in \{0, \ldots, k\}$ tal que $\hat{x}_i \in B$ y $\hat{x}_{i+1} \in C$, de tal forma que, por el lema 1.2, $[\hat{x}_i, \hat{x}_{i+1}] \nsubseteq B \cup C = A$.

Resumiendo, si A no fuera conexo, no existiría una poligonal contenida en A que una a $\hat{x}$ y $\hat{y}$, lo que contradice nuestra hipótesis. ■

Volviendo al recíproco de la proposición 1.12, aún cuando geométricamente es muy sencillo *mostrar* un ejemplo de un conjunto que sea conexo pero no convexo (ver figura 1.16), con base en el resultado anterior podemos *probar* que el siguiente ejemplo ilustra el mismo hecho.

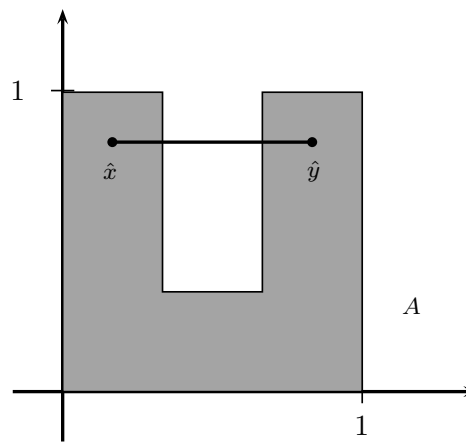


Figura 1.16: El conjunto A es conexo pero no es convexo, ya que el segmento $[\hat{x}, \hat{y}]$ no está contenido en A.

Ejemplo 1.11 *Sea*

$$A = \mathbb{R}^2 \setminus \{(x, y) \in \mathbb{R}^2 \mid x \leq 0 \ y \ y = 0\}.$$

El lector probará en el problema 16 que este conjunto es abierto. Aquí sólo mostraremos que dado cualquier $\hat{x} = (x, y) \in A$, el segmento que lo une con el punto $\hat{x}_0 = (1, 0) \in A$ está totalmente contenido en A (ver figura 1.17), lo que mostrará que este conjunto satisface la condición de suficiencia de la proposición anterior y por lo tanto será conexo.

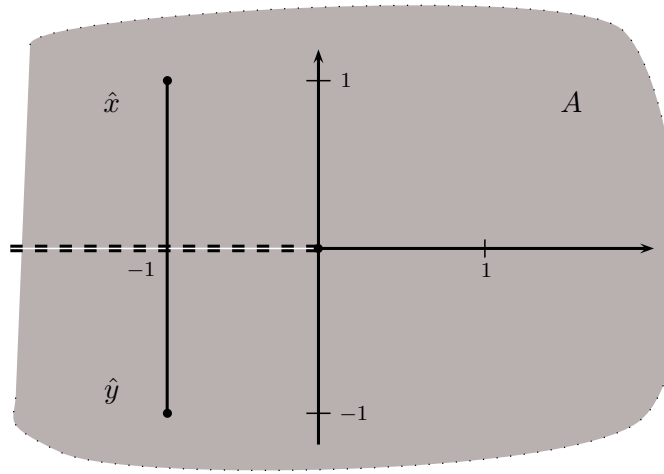


Figura 1.17: El conjunto $A = \mathbb{R}^2 \setminus \{(x, y) \in \mathbb{R}^2 \mid x \leq 0 \text{ y } y = 0\}$ es conexo pero no es convexo, ya que el segmento $[\hat{x}, \hat{y}]$ no está contenido en A.

Dado $(x, y) \in A$, distinguimos tres casos:

1. *$y = 0$ y $x > 0$. En este caso, el segmento $[\hat{x}, \hat{x}_0] = [(x, 0), (1, 0)]$ está dado por*

$$\{(x + t(1 - x), 0) = ((1 - t)x + t, 0) \in \mathbb{R}^2 \mid t \in [0, 1]\}$$

y como $(1-t)x+t > 0$ para toda $t \in [0, 1]$, se tiene que está contenido en A.

2. *$y > 0$ y $x \in \mathbb{R}$. En este caso, el segmento $[\hat{x}, \hat{x}_0] = [(x, y), (1, 0)]$ está dado por*

$$\{(x + t(1 - x), y - ty) = ((1 - t)x + t, (1 - t)y) \in \mathbb{R}^2 \mid t \in [0, 1]\}$$

de tal forma que, como $(1 - t)y > 0$ para toda $t \in [0, 1)$ y para $t = 1$ obtenemos el punto $(1, 0) = \hat{x}_0$, entonces en este caso también dicho segmento está contenido en A.

3. *$y < 0$ y $x \in \mathbb{R}$. En este caso el segmento $[\hat{x}, \hat{x}_0] = [(x, y), (1, 0)]$ está dado por el mismo conjunto del inciso anterior, sólo que ahora $(1-t)y < 0$ para toda $t \in [0, 1)$ y para $t = 1$ obtenemos otra vez el punto $(1, 0) = \hat{x}_0$, de modo que nuevamente el segmento estará contenido en A.*

Ahora sólo resta probar que A no es convexo. En efecto, si tomamos los puntos $\hat{x} = (-1, 1)$ y $\hat{y} = (-1, -1)$ se tiene que $\hat{x}, \hat{y} \in A$ y sin embargo el punto $(-1, 0) = \hat{x} + (1/2)(\hat{y} - \hat{x}) \in [\hat{x}, \hat{y}]$ no pertenece a A, es decir, $[\hat{x}, \hat{y}] \nsubseteq A$.

Concluimos esta sección con un resultado que nos permite caracterizar a los conjuntos disconexos (y por tanto a los conexos) en términos de conjuntos abiertos. Aún cuando no es la caracterización más conocida en términos de este tipo de conjuntos (la cual podremos probar hasta el capítulo 2), este caso también resultará ser de gran utilidad.

Proposición 1.14 *Sea $A \subset \mathbb{R}^n$. A es disconexo si y sólo si existen $U, V \subset \mathbb{R}^n$ conjuntos abiertos tales que:*

1. *$A \subset U \cup V$,*
2. *$A \cap U \neq \varnothing$ y $A \cap V \neq \varnothing$, y*
3. *$A \cap U \cap V = \varnothing$.*

Demostración. Si A es disconexo, sabemos que existen $B, C \subset \mathbb{R}^n$ no vacíos y separados tales que $A = B \cup C$. Sea $U = \mathbb{R}^n \setminus \bar{C}$ y $V = \mathbb{R}^n \setminus \bar{B}$; nótese que U y V son abiertos puesto que cada uno de ellos es el complemento de un conjunto cerrado, razón por la cual sólo nos restará probar que satisfacen las propiedades requeridas.

Como B y C están separados, sabemos que $\bar{B} \cap C = \varnothing = B \cap \bar{C}$, de tal forma que $C \subset \mathbb{R}^n \setminus \bar{B} = V$ y $B \subset \mathbb{R}^n \setminus \bar{C} = U$. De aquí que

$$\begin{aligned} A &= B \cup C \\ &\subset (\mathbb{R}^n \setminus \bar{C}) \cup (\mathbb{R}^n \setminus \bar{B}) \\ &= U \cup V. \end{aligned}$$

Por otra parte,

$$\begin{aligned} A \cap U &= A \cap (\mathbb{R}^n \setminus \bar{C}) \\ &\supset A \cap B \\ &= B \\ &\neq \varnothing, \end{aligned}$$

y análogamente

$$\begin{aligned} A \cap V &= A \cap (\mathbb{R}^n \setminus \bar{B}) \\ &\supset A \cap C \\ &= C \\ &\neq \varnothing. \end{aligned}$$

Finalmente,

$$\begin{aligned} A \cap U \cap V &= A \cap (\mathbb{R}^n \setminus \bar{B}) \cap (\mathbb{R}^n \setminus \bar{C}) \\ &= A \cap (\mathbb{R}^n \setminus (\bar{B} \cup \bar{C})) \\ &= (B \cup C) \cap (\mathbb{R}^n \setminus (\bar{B} \cup \bar{C})) \\ &= \varnothing. \end{aligned}$$

Para probar la implicación recíproca, tomamos $B = A \cap U$ y $C = A \cap V$; por los incisos 1 y 2 de la hipótesis, tenemos que $B \neq \varnothing, C \neq \varnothing$ y que $A = B \cup C$, de modo que sólo nos restaría mostrar que estos conjuntos están separados.

Supongamos que

$$\bar{B} \cap C = \overline{(A \cap U)} \cap (A \cap V) \neq \varnothing.$$

Si $\hat{x} \in \overline{(A \cap U)} \cap (A \cap V) \subset V$, como V es abierto, existe $r > 0$ tal que $B_r(\hat{x}) \subset V$. Por otra parte, como $\hat{x} \in \overline{(A \cap U)}$, por el inciso (a) del problema 27 sabemos que

$$\varnothing \neq B_r(\hat{x}) \cap (A \cap U) \subset V \cap (A \cap U) = A \cap U \cap V.$$

lo que contradice el inciso 3 de la hipótesis. Por tanto $\bar{B} \cap C = \overline{(A \cap U)} \cap (A \cap V) = \varnothing$. Análogamente se prueba que $\bar{C} \cap B = \overline{(A \cap V)} \cap (A \cap U) = \varnothing$. ■

Como mencionamos anteriormente, como un corolario inmediato tenemos la siguiente caracterización de los conjuntos conexos.

Corolario 1.2 *Sea $A \subset \mathbb{R}^n$. A es conexo si y sólo si no existen $U, V \subset \mathbb{R}^n$ conjuntos abiertos tales que:*

1. $A \subset U \cup V$,
2. $A \cap U \neq \varnothing$ *y* $A \cap V \neq \varnothing$, *y*
3. $A \cap U \cap V = \varnothing$.

1.6. Otros sistemas coordenados

Al inicio de este capítulo mostramos algunas "situaciones" que se podían describir por medio de funciones cuyas variables (independiente y/o dependiente) eran una posición o una flecha. En todos estos ejemplos usamos el hecho de que una posición o una flecha (en el plano o en el espacio) se puede representar, una vez que se ha establecido un sistema coordenado cartesiano, por una pareja o una terna de números reales. Cabe recordar que en estos casos, la función involucrada se tendría que expresar en términos de estas coordenadas.

Hay "situaciones" en las que para representar una posición o una flecha es conveniente usar ciertas cantidades. Un ejemplo de estas cantidades, para el caso de posiciones o flechas en el plano, serían la distancia a un punto fijo y el ángulo con respecto a una recta fija. Estas cantidades dan lugar a otros sistemas coordenados, que son los que nos proponemos estudiar en esta sección.

Aún cuando no siempre es necesario, estos otros sistemas coordenados se pueden construir con base en un sistema cartesiano establecido previamente, que es justo como lo haremos en este texto.

1.6.1. Coordenadas polares

Para el caso del plano, si $\hat{x}$ representa un punto o una flecha, además de asignarle sus coordenadas (x_0, y_0) en un sistema cartesiano dado, podemos asignarle otra pareja de números (ρ, θ), a la que llamaremos *unas coordenadas polares* de $\hat{x}$, en donde ρ es igual a la distancia que hay de $\hat{x}$ al origen (es decir, la magnitud o norma de $\hat{x}$), y θ es el ángulo (dirigido) formado por la parte positiva del eje X y la semirrecta que parte del origen y pasa por $\hat{x}$ (ver figura 1.18).

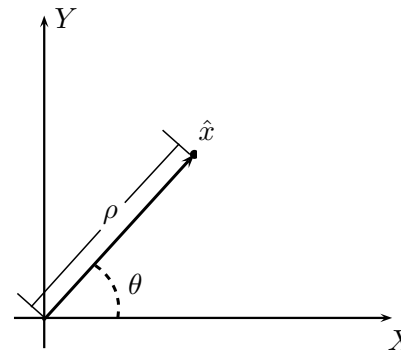


Figura 1.18: Obtención de coordenadas polares de un punto (o una flecha) $\hat{x}$ del plano.

Como el lector recordará, la interpretación geométrica de las coordenadas cartesianas (x_0, y_0) de un punto $\hat{x}$ en $\mathbb{R}^2$, consiste en ver a $\hat{x}$ como el único punto en el que se intersectan las rectas determinadas por las ecuaciones $x = x_0$ (paralela al eje Y, y que es justo todo el conjunto de puntos del plano cuya coordenada cartesiana x es igual a x_0) y $y = y_0$ (paralela al eje X, y que es justo todo el conjunto de puntos del plano cuya coordenada cartesiana y es igual a y_0).

En el caso de las coordenadas polares hay una interpretación geométrica análoga. En efecto, como el lector se podrá convencer fácilmente, dado un punto $\hat{x}$ en $\mathbb{R}^2$, distinto del origen $\hat{0}$, existen una única circunferencia con centro en el origen y una única semirrecta que parte del origen, cuya

intersección es el punto $\hat{x}$. Pues bien, si (ρ_0, θ_0) son las coordenadas polares de este punto, ρ_0 proporciona el radio de la circunferencia (que es justo el conjunto de puntos del plano cuya coordenada polar ρ es igual a la constante ρ_0), y el ángulo θ_0, medido con respecto a la parte positica del eje X, determina a la semirrecta (que es justo todo el conjunto de puntos del plano cuya coordenada polar θ es igual a la constante θ_0), que tienen como intersección al punto (o vector) $\hat{x}$ (ver figura 1.19).

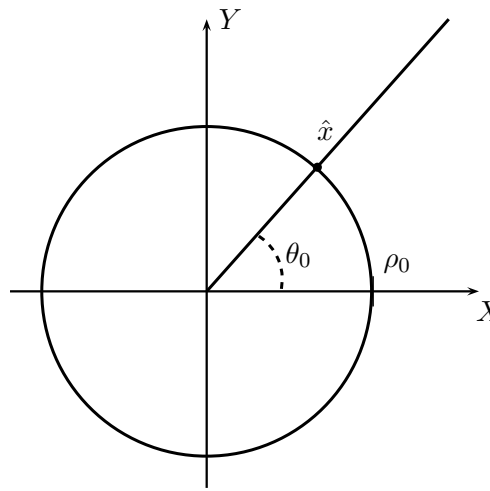


Figura 1.19: Las coordenadas polares (ρ_0, θ_0) nos proporcionan el radio (ρ_0) de la única circunferencia, y el ángulo (θ_0) de la única semirrecta que parte del origen, que tienen como intersección al punto (o vector) $\hat{x}$.

Otro aspecto importante que hay que hacer notar con relación a estas nuevas coordenadas es que para asignarlas no es necesario recurrir a todas las parejas ordenadas de $\mathbb{R}^2$; es decir, basta con tomar las parejas (ρ, θ) tales que $0 \leq \rho$ y $0 \leq \theta < 2\pi$ para que todo punto (o flecha) del plano tenga asociada una de ellas.

De hecho, salvo por el origen $\hat{0}$ (cuyas coordenadas polares pueden estar dadas por cualquier pareja de la forma $(0, \theta)$, con θ cualquier número real), la asignación de unas coordenadas polares establece una biyección entre

un plano menos un punto (aquel que se haya elegido como el origen) y el subconjunto $(0, \infty) \times [0, 2\pi) \subset \mathbb{R}^2$. Más aún, nótese que el ángulo siempre se puede elegir en cualquier intervalo de la forma $\theta_0 \leq \theta < \theta_0 + 2\pi$, en donde θ_0 es un ángulo fijo.

De esta forma, un mismo punto (o vector) $\hat{x}$ tiene muchas coordenadas polares, razón por la cual al inicio de esta subsección hablamos de *unas* (y no de *las*) coordenadas polares del punto $\hat{x}$.

En reciprocidad con lo anterior, también es importante hacer notar que cualquier pareja $(\rho, \theta) \in \mathbb{R}^2$ (incluso con $\rho < 0$) se puede "interpretar" como coordenadas polares de un punto (o vector) $\hat{x}$ del plano, obteniendo (o "construyendo") $\hat{x}$ de la siguiente manera: dada una pareja $(\rho, \theta) \in \mathbb{R}^2$, hágase una rotación del eje X por θ radianes, y sobre ese eje rotado, localice el número ρ; este será el punto $\hat{x}$ que le corresponda a la pareja (ρ, θ) (ver figura 1.20).

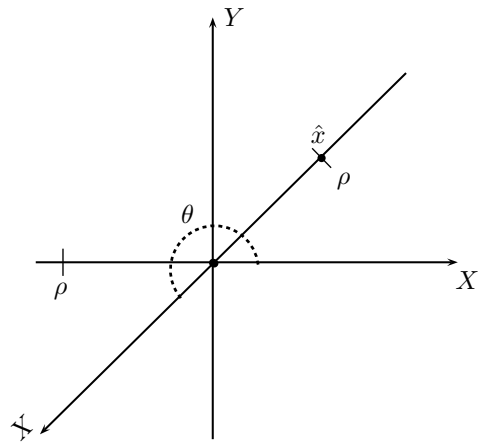


Figura 1.20: Cualquier pareja $(\rho, \theta) \in \mathbb{R}^2$ (en este caso con $\rho < 0$) se puede "interpretar" como coordenadas polares de un punto (o vector) $\hat{x}$ del plano, el cual se obtiene de la siguiente manera: hágase una rotación del eje X por θ radianes, y sobre ese eje rotado, localice el número ρ; éste será el punto $\hat{x}$ que le corresponda a la pareja (ρ, θ).

Con base en lo anterior, nótese que podemos concluir que, si (ρ, θ) y (ρ', θ') son coordenadas polares de un mismo vector $\hat{x}$ del plano, entonces se debe cumplir que $\rho = \pm\rho'$ y $\theta = \theta' + k\pi$ para alguna $k \in \mathbb{Z}$. Más específicamente, se tiene que $\rho = \rho'$ si y sólo si $\theta = \theta' + 2k\pi$ para alguna $k \in \mathbb{Z}$, y $\rho = -\rho'$ si y sólo si $\theta = \theta' + (2k+1)\pi$ para alguna $k \in \mathbb{Z}$.

Hay otra observación importante con relación a las coordenadas polares: las operaciones de suma y producto por un escalar que definimos para las parejas de $\mathbb{R}^2$, ya no se corresponden con la suma y producto por un escalar que definimos geométricamente para vectores (o flechas) del plano. Es decir, si (ρ_1, θ_1) y (ρ_2, θ_2) son coordenadas polares de $\hat{x}_1$ y $\hat{x}_2$, respectivamente, entonces la pareja $(\rho_1 + \rho_2, \theta_1 + \theta_2)$ no son necesariamente coordenadas polares de $\hat{x}_1 + \hat{x}_2$ (en donde esta última suma de vectores es la que se obtiene por medio de la ley del paralelogramo); y si $\lambda \in \mathbb{R}$, entonces la pareja $(\lambda\rho_1, \lambda\theta_1)$ no son necesariamente coordenadas polares de $\lambda\hat{x}_1$. Sin duda un problema interesante consiste en encontrar coordenadas polares para $\hat{x}_1 + \hat{x}_2$ y $\lambda\hat{x}_1$ en términos de coordenadas polares de $\hat{x}_1$ y $\hat{x}_2$, y por lo mismo lo dejamos como un problema para el lector.

Lo que sí queremos mencionar aquí es la interpretación geométrica de ciertas "operaciones aritméticas" con coordenadas polares; específicamente, si (ρ, θ) son coordenadas polares de un vector $\hat{x}$ y $h \in \mathbb{R}$, entonces $(\rho, \theta + h)$ son coordenadas polares de un vector $\hat{x}'_h$ que se obtiene rotando h radianes a $\hat{x}$; y $(\rho + h, \theta)$ son coordenadas polares de un vector $\hat{x}_h$ que está en la misma dirección que está $\hat{x}$, sólo modificando su magnitud por una cierta cantidad h (y suponiendo que $\rho + h$ no tiene signo diferente a ρ; ¿qué sucede geométricamente si ρ y $\rho + h$ tienen signo diferente?) (ver figura 1.21).

Finalmente, con base en las funciones trigonométricas, deducimos las ecuaciones que nos permiten obtener, dadas cualesquiera coordenadas polares (ρ, θ) de un vector $\hat{x}$, sus correspondientes coordenadas cartesianas (x, y) (ambas consideradas sobre el mismo sistema coordenado cartesiano XY). Estas ecuaciones son las siguientes:

$$\begin{aligned} x &= \rho\cos(\theta) \\ y &= \rho\,\mathrm{sen}(\theta) \end{aligned} \tag{1.12}$$

(ver figura 1.22).

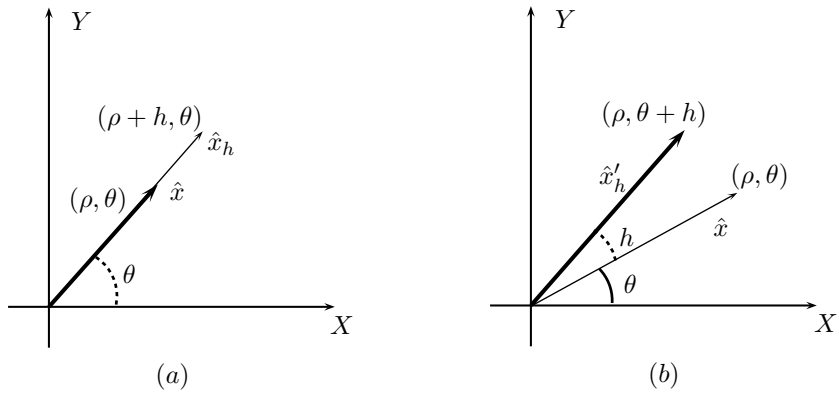


Figura 1.21: Si un vector $\hat{x}$ tiene coordenadas polares (ρ,θ), entonces el vector $\hat{x}_h$ de coordenadas polares $(\rho+h,\theta)$ está en la misma dirección que $\hat{x}$ (si ρ y $\rho+h$ tienen el mismo signo) (a), y el vector $\hat{x}'_h$ de coordenadas polares $(\rho,\theta+h)$ se obtiene rotando h radianes el vector $\hat{x}$ (b).

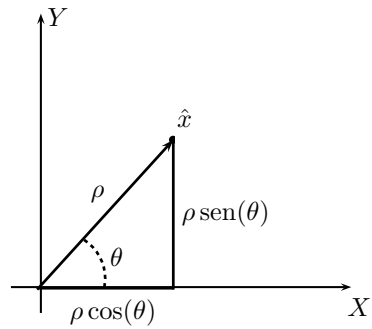


Figura 1.22: Dadas cualesquiera coordenadas polares (ρ,θ) de un vector $\hat{x}$, sus correspondientes coordenadas cartesianas son $(\rho\cos(\theta),\rho\,\mathrm{sen}(\theta))$.

Recíprocamente, si conocemos las coordenadas cartesianas (x,y) de un vector $\hat{x}$, podemos obtener unas coordenadas polares de $\hat{x}$ de la siguiente forma: dado que ρ representa la distancia de $\hat{x}$ al origen (es decir, la norma de $\hat{x}$), sabemos que

$$\rho = \sqrt{x^2 + y^2}$$

(lo que también se puede confirmar a partir de las dos identidades anteriores).

Obtener una expresión para θ es un poco más elaborado y hay que analizar algunos casos. Como ya habíamos mencionado, si $\hat{x} = \hat{0}$, cualquier pareja de la forma $(0, \theta)$ ($\theta \in \mathbb{R}$) son coordenadas polares de $\hat{x}$. Si $x = 0$ y $y > 0$ (es decir que $\hat{x}$ está en la parte positiva del eje Y), bastará con tomar $\theta = \pi/2$ o en general

$$\theta = \frac{\pi}{2} + 2k\pi.$$

Y si $x = 0$ y $y < 0$ (es decir que $\hat{x}$ está en la parte negativa del eje Y), bastará con tomar $\theta = 3\pi/2$ o en general

$$\theta = \frac{3\pi}{2} + 2k\pi,$$

con $k \in \mathbb{Z}$.

Si $x \neq 0$, de las identidades 1.12 se tiene que

$$\begin{aligned} \frac{y}{x} &= \frac{\operatorname{sen}(\theta)}{\cos(\theta)} \\ &= \tan(\theta), \end{aligned}$$

de tal forma que si arctan es la rama de la función inversa de la función tangente que toma sus valores entre $-\pi/2$ y $\pi/2$, entonces, cuando se tenga $x > 0$, bastará con tomar $\theta = \arctan(y/x)$, o en general

$$\theta = \arctan\left(\frac{y}{x}\right) + 2k\pi$$

con $k \in \mathbb{Z}$; y si $x < 0$, bastará con tomar $\theta = \arctan(y/x) + \pi$ o en general

$$\theta = \arctan\left(\frac{y}{x}\right) + (2k+1)\pi,$$

con $k \in \mathbb{Z}$.

1.6.2. Coordenadas cilíndricas

Para el caso de puntos y/o flechas en el espacio, y nuevamente partiendo de un sistema cartesiano dado, vamos a dar otra forma de asignarles ternas de números que los representen, de la siguiente manera.

Si $\hat{x}$ es un punto o una flecha en el espacio cuyas coordenadas cartesianas (en el sistema XYZ dado) son (x_0, y_0, z_0), nos fijamos en el vector del plano XY que tiene coordenadas cartesianas (x_0, y_0) (y que geométricamente se obtiene de "proyectar" a $\hat{x}$ en el plano XY). Si ahora la pareja (ρ_0, θ_0) son unas coordenadas polares del vector (x_0, y_0), decimos entonces que la terna (ρ_0, θ_0, z_0) son *unas coordenadas cilíndricas* del vector $\hat{x}$ (ver figura 1.23).

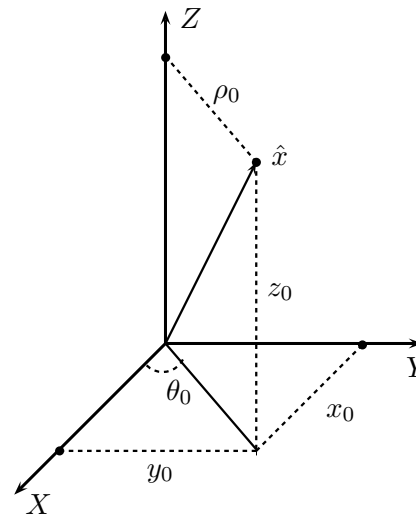


Figura 1.23: Si la pareja (ρ_0, θ_0) son unas coordenadas polares del vector (x_0, y_0) del plano XY, decimos entonces que la terna (ρ_0, θ_0, z_0) son *unas coordenadas cilíndricas* del vector $\hat{x}$.

Como en el caso de las coordenadas polares, las coordenadas cilíndricas también se pueden interpretar geométricamente. En efecto, dado un punto $\hat{x}$ en $\mathbb{R}^3$, también distinto del origen $\hat{0}$, el lector se podrá convencer fácilmente de que existen: un único cilindro circular recto cuyo eje sea el eje Z (¡que

puede coincidir con este eje!); un único semiplano que tiene como "borde" al eje Z (y por tanto perpendicular al plano XY); y un único plano paralelo al plano XY, tales que su intersección sólo sea el punto $\hat{x}$ (ver figura 1.24). Pues bien, si (ρ_0, θ_0, z_0) son las coordenadas cilíndricas de este punto, ρ_0 proporciona el radio de este cilindro (que es justo el conjunto de puntos del espacio cuya coordenada cilíndrica ρ es igual a la constante ρ_0), el ángulo θ_0, medido con respecto a la parte positica del eje X, determina al semiplano (que es justo todo el conjunto de puntos del espacio cuya coordenada cilíndrica θ es igual a la constante θ_0), y z_0 da la "altura" del plano paralelo al plano XY (que es justo todo el conjunto de puntos del espacio cuya coordenada cilíndrica z es igual a la constante z_0), que tienen como intersección al punto (o vector) $\hat{x}$.

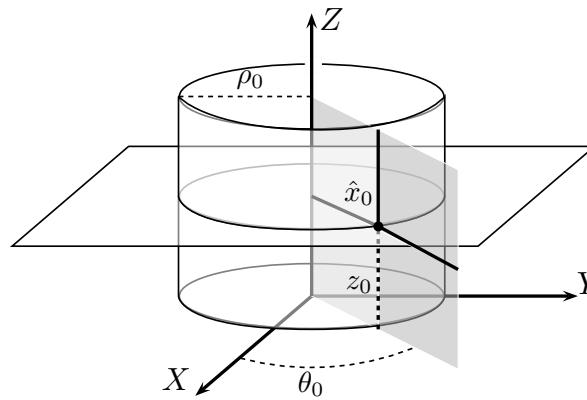


Figura 1.24: Las coordenadas cilíndricas (ρ_0, θ_0, z_0) de un punto $\hat{x}$ nos proporcionan el radio ρ_0 del cilindro circular recto, el ángulo θ_0 del semiplano "abisagrado" al eje Z y perpendicular al plano XY, y la altura z_0 del plano paralelo al plano XY, que tienen como intersección únicamente al punto $\hat{x}$.

Nuevamente es importante hacer notar que para asociar unas coordenadas cilíndricas a un vector $\hat{x}$, no es necesario recurrir a todas las ternas de $\mathbb{R}^3$, pues basta con que $0 \leq \rho$, $0 \leq \theta < 2\pi$ y $z \in \mathbb{R}$. Es decir, a todo vector en el espacio se le puede asignar unas coordenadas cilíndricas de tal forma que éstas

estén en el subconjunto $[0,\infty)\times[0,2\pi)\times(-\infty,\infty) \subset \mathbb{R}^3$, e incluso, de manera más general, en un subconjunto de la forma $[0,\infty)\times[\theta_0,\theta_0+2\pi)\times(-\infty,\infty) \subset \mathbb{R}^3$, con $\theta_0 \in \mathbb{R}$ fijo.

No obstante, como en el caso de las coordenadas polares, cualquier terna $(\rho,\theta,z) \in \mathbb{R}^3$ se puede interpretar como coordenadas cilíndricas de un vector del espacio. Para hacer esto, basta con localizar en el plano XY el vector que se le debe asignar a la pareja (ρ,θ) (de acuerdo con el procedimiento descrito en la sección anterior) y después "elevarlo" a la altura z (ver figura 1.25).

Con base en lo anterior tenemos que, si dos ternas (ρ,θ,z) y (ρ',θ',z') son coordenadas cilíndricas del mismo vector $\hat{x}$, entonces podemos asegurar que $\rho = \pm\rho'$, $\theta = \theta' + k\pi$ para alguna $k \in \mathbb{Z}$, y $z = z'$.

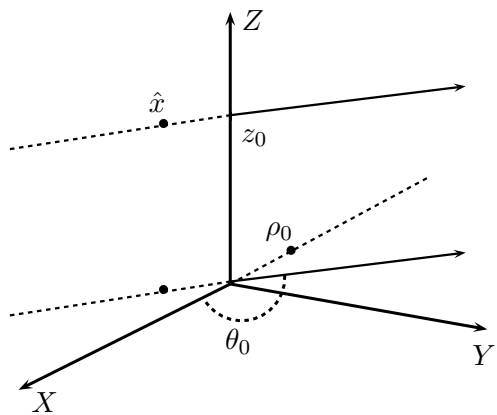


Figura 1.25: Cualquier terna $(\rho_0,\theta_0,z_0) \in \mathbb{R}^3$ se puede interpretar como coordenadas cilíndricas de un punto $\hat{x}$ del espacio.

Análogamente a lo que sucede con las coordenadas polares, la "aritmética" definida entre ternas no se corresponde con la "aritmética" definida geométricamente entre vectores del espacio. Sin embargo, también es importante identificar geométricamente el efecto de sumar una cierta cantidad $h \in \mathbb{R}$ en una sóla de estas coordenadas.

Si (ρ_0, θ_0, z_0) son coordenadas cilíndricas de un vector $\hat{x}$, entonces $(\rho_0 + h, \theta_0, z_0)$ son coordenadas cilíndricas del vector que está en el mismo plano que pasa por $\hat{x}$ y que contiene al eje Z (si suponemos que $\rho_0 + h$ no tiene signo diferente a ρ_0; ¿qué sucede geométricamente si ρ_0 y $\rho_0 + h$ tienen signo diferente?), a la misma "altura" sobre el eje Z, pero con la diferencia de que su proyección sobre el plano XY es un vector de norma $\rho_0 + h$ ($|\rho_0 + h|$ en el caso general) (ver figura 1.26 (a)).

Si ahora consideramos $(\rho_0, \theta_0 + h, z_0)$, éstas serán coordenadas cilíndricas del vector que se obtiene al rotar, con respecto el eje Z, h radianes el vector $\hat{x}$ (ver figura 1.26 (b)).

Finalmente, si ahora tomamos $(\rho_0, \theta_0, z_0 + h)$, éstas serán coordenadas cilíndricas del vector que se obtiene "cambiando" la punta del vector $\hat{x}$ a la altura $z_0 + h$ (ver figura 1.26 (c)); en particular, este vector y el vector $\hat{x}$ tendrán la misma proyección sobre el plano XY.

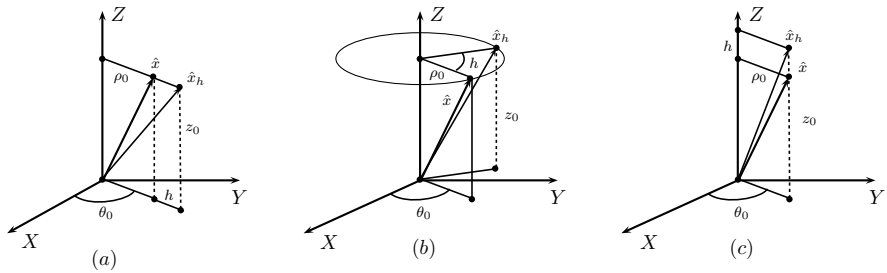


Figura 1.26: Si el vector $\hat{x}$ tiene coodenadas cilíndricas (ρ_0, θ_0, z_0), entonces en los siguientes casos el vector $\hat{x}_h$ tiene coordenadas cilíndricas: (a) $(\rho_0 + h, \theta_0, z_0)$; (b) $(\rho_0, \theta_0 + h, z_0)$; (c) $(\rho_0, \theta_0, z_0 + h)$.

Como en el caso de las coordenadas polares, si (ρ, θ, z) son cualesquiera coordenadas cilíndricas de un vector $\hat{x}$, usando nuevamente las funciones trigonométricas, obtenemos que la terna (x, y, z), con

$$\begin{aligned} x &= \rho \cos(\theta) \\ y &= \rho \operatorname{sen}(\theta) \\ z &= z \end{aligned}$$

son las coordenadas cartesianas de $\hat{x}$. Nótese que la tercera (y extraña) identidad anterior refleja el abuso de notación que cometemos al nombrar con la misma letra a la tercera coordenada cilíndrica y a la tercera coordenada cartesiana del vector $\hat{x}$, lo cual se "justifica" por el hecho de que ambas tienen el mismo significado geométrico.

Recíprocamente, para obtener unas coordenadas cilíndricas de un vector $\hat{x}$ a partir de sus coordenadas cartesianas (x, y, z), calculamos ρ y θ procediendo justo igual que en el caso de las coordenadas polares, y la coordenada polar z se toma igual a la coordenada cartesiana z (por las razones que ya explicamos en el párrafo anterior).

1.6.3. Coordenadas esféricas

Concluimos esta breve sección introduciendo una forma más de asignarle una terna de números reales a un vector $\hat{x}$ del espacio.

Como en el caso anterior, partiendo de un sistema cartesiano XYZ dado, asignamos a $\hat{x}$ la terna de números reales (ρ, θ, φ), de la siguiente forma: ρ (como en el caso de las coordenadas polares) representa la magnitud de $\hat{x}$, θ es el ángulo (dirigido) formado por la parte positiva del eje X y la proyección del vector $\hat{x}$ sobre el plano XY, siempre que esta proyección sea diferente del vector $\hat{0}$; si dicha proyección es el vector $\hat{0}$, hacemos $\theta = 0$; y finalmente, φ será el ángulo (dirigido) formado por la parte positiva del eje Z y el vector $\hat{x}$ ([6]) (ver figura 1.27). A la terna (ρ, θ, φ) construida de esta forma le llamaremos *unas coordenadas esféricas* del vector $\hat{x}$.

Nuevamente, las coordenadas esféricas se pueden interpretar geométricamente. En este caso, dado $\hat{x}$ en $\mathbb{R}^3$, también distinto del origen $\hat{0}$, se tiene que existen: una única esfera con centro en el origen; un único semiplano que tiene como "borde" al eje Z (y por tanto perpendicular al plano XY); y un único cono circular con vértice en el origen, tales que su intersección sólo sea el punto $\hat{x}$ (ver figura 1.28). De esta forma, si $(\rho_0, \theta_0, \varphi_0)$ son las coordenadas esféricas de este punto, ρ_0 proporciona el radio de la esfera (que es justo el conjunto de puntos del espacio cuya coordenada esférica ρ

[6] En algunos textos, el ángulo φ se toma como el ángulo (dirigido) formado por el plano XY y el vector $\hat{x}$.

es igual a la constante ρ_0), el ángulo θ_0, medido con respecto a la parte positica del eje X, determina al semiplano (que es justo todo el conjunto de puntos del espacio cuya coordenada esférica θ es igual a la constante θ_0), y φ_0 determina al cono (que es justo todo el conjunto de puntos del espacio cuya coordenada esférica φ es igual a la constante φ_0), que tienen como intersección al punto (o vector) $\hat{x}$.

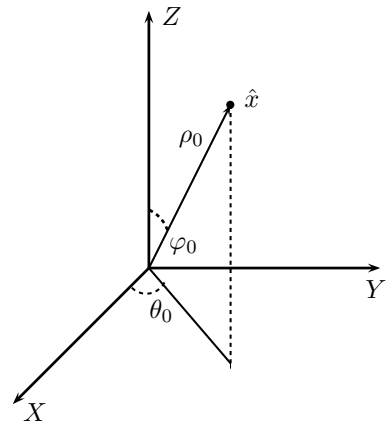


Figura 1.27: La terna $(\rho_0, \theta_0, \varphi_0)$ son *unas coordenadas esféricas* del vector $\hat{x}$.

De la misma forma que en los dos casos anteriores, para asignar unas coordenadas esféricas a un vector $\hat{x}$ no es necesario recurrir a todas las ternas de $\mathbb{R}^3$; como se podrá notar, basta con que $0 \leq \rho$, $0 \leq \theta < 2\pi$ y $0 \leq \varphi \leq \pi$, o de manera más general, que $0 \leq \rho$, $\theta_0 \leq \theta < \theta_0 + 2\pi$ y $\varphi_0 \leq \varphi \leq \varphi_0 + \pi$, con $\theta_0, \varphi_0 \in \mathbb{R}$ fijos.

También, como en los dos casos anteriores, toda terna $(\rho, \theta, \varphi) \in \mathbb{R}^3$ se puede "interpretar" como coordenadas esféricas de un vector $\hat{x}$ del espacio, de la siguiente forma: sobre el plano XZ, y con respecto a la parte positiva del eje Z, rótese φ radianes el eje Z; a continuación, tomando como eje de rotación el eje Z, y con respecto a la parte positiva del eje X, rote la recta real resultante (o si prefiere todo el plano XZ) θ radianes; finalmente, sobre esta última recta real que se obtiene, ubique el número ρ.

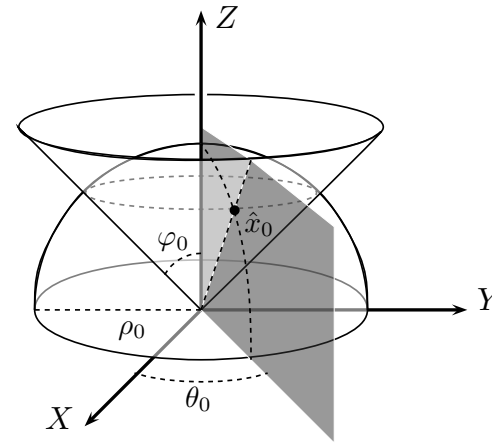


Figura 1.28: Las coordenadas esféricas $(\rho_0, \theta_0, \varphi_0)$ de un punto $\hat{x}_0$ nos proporcionan el radio ρ_0 de la esfera, el ángulo θ_0 del semiplano perpendicular al plano XY, y el ángulo φ_0 del cono que tienen como intersección únicamente al punto $\hat{x}_0$.

El vector $\hat{x}$ determinado por el procedimiento anterior será el vector que asociaremos a la terna (ρ, θ, φ) y diremos que los elementos de esta terna también son coordenadas esféricas de $\hat{x}$. Todo lo anterior sin duda requiere de un

Ejemplo 1.12 *Sea $\hat{x}$ el vector que en un sistema cartesiano dado tiene coordenadas (cartesianas)* $(1, 1, 1)$. *De acuerdo con el procedimiento de asignación de coordenadas esféricas, tenemos que la terna* $(\sqrt{3}, \pi/4, \pi/4)$ *nos da unas coordenadas de este tipo para $\hat{x}$.*

Por otra parte, y de acuerdo con el método de asignación de un vector a una terna, nótese que las ternas $(\sqrt{3}, -3\pi/4, -\pi/4)$, $(-\sqrt{3}, -3\pi/4, 3\pi/4)$, *y* $(-\sqrt{3}, \pi/4, -3\pi/4)$ *también son coordenadas esféricas de nuestro vector $\hat{x}$ (ver figura 1.29).*

Aun cuando existen diferentes coordenadas esféricas para un mismo vector $\hat{x}$ (como sucede con los otros tipos de coordenadas que hemos visto en

esta sección), si (ρ, θ, φ) y $(\rho', \theta', \varphi')$ son dos de ellas, lo que podemos asegurar es que se debe tener que $\rho = \pm\rho'$, que $\theta = \theta' + k\pi$ para alguna $k \in \mathbb{Z}$, y que $\varphi = \varphi' + k'(\pi/2)$ también para alguna $k' \in \mathbb{Z}$.

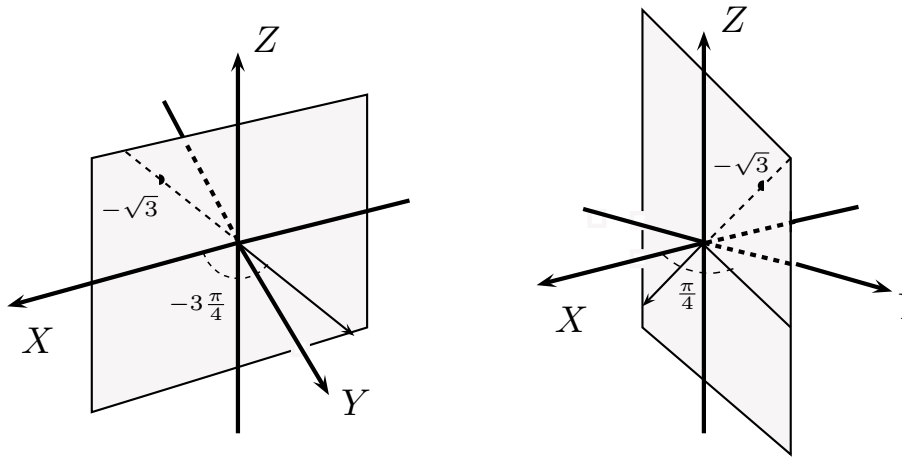


Figura 1.29: El punto $\hat{x}$, el cual también tiene como coodenadas esféricas a la terna $(-\sqrt{3}, \pi/4, -3\pi/4)$, se obtiene de la siguiente forma: sobre el plano XZ, y con respecto a la parte positiva del eje Z, rote $-\frac{3\pi}{4}$ radianes el eje Z; a continuación, tomando como eje de rotación el eje Z, y con respecto a la parte positiva del eje X, rote la recta real resultante $\frac{\pi}{4}$ radianes; finalmente, sobre esta última recta real que se obtiene, ubique el número $-\sqrt{3}$.

Como en los casos anteriores, la "aritmética" definida entre ternas no se corresponde con la "aritmética" definida geométricamente entre vectores del espacio. Sin embargo, como también hicimos en los otros casos, es importante identificar geométricamente el efecto de sumar una cierta cantidad $h \in \mathbb{R}$ en una sóla de estas coordenadas.

De esta forma, si (ρ, θ, φ) son coordenadas esféricas de un vector $\hat{x}$, entonces el vector que tenga coordenadas esféricas $(\rho + h, \theta, \varphi)$ será aquel que está en la misma recta que tiene a $\hat{x}$, con norma $|\rho + h|$ y apuntando en la misma dirección que (o contraria a) $\hat{x}$, dependiendo de si ρ y $\rho + h$ tienen el mismo signo (o signo diferente); asimismo, el vector que tenga coordenadas

esféricas $(\rho, \theta + h, \varphi)$ será aquel que se obtiene de rotar h radianes al vector $\hat{x}$, tomando como eje de rotación al eje Z; finalmente, el vector que tenga coordenadas esféricas $(\rho, \theta, \varphi + h)$ será aquel que se obtiene de rotar h radianes al vector $\hat{x}$, realizando dicha rotación sobre el plano que contenga a $\hat{x}$ y al eje Z (ver figura 1.30).

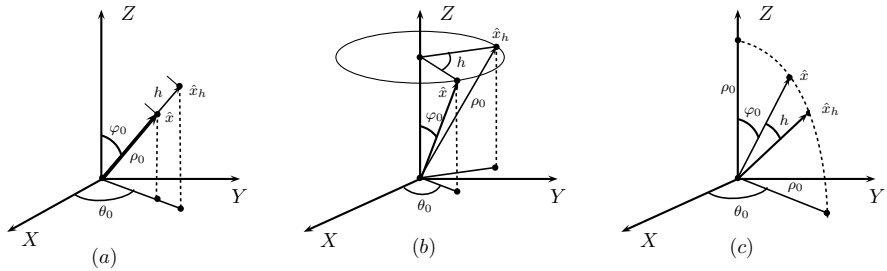


Figura 1.30: Si el vector $\hat{x}$ tiene coodenadas esféricas $(\rho_0, \theta_0, \varphi_0)$, entonces en los siguientes casos el vector $\hat{x}_h$ tiene coordenadas esféricas: (a) $(\rho_0 + h, \theta_0, \varphi_0)$; (b) $(\rho_0, \theta_0 + h, \varphi_0)$ y (c) $(\rho_0, \theta_0, \varphi_0 + h)$.

Finalmente, recurriendo de nuevo a las funciones trigonométricas, si (ρ, θ, φ) son cualesquiera coordenadas esféricas de un vector $\hat{x}$, obtenemos que la terna (x, y, z), con

$$\begin{aligned} x &= \rho \operatorname{sen}(\varphi) \cos(\theta) \\ y &= \rho \operatorname{sen}(\varphi) \operatorname{sen}(\theta) \\ z &= \rho \cos(\varphi) \end{aligned} \tag{1.13}$$

son las coordenadas cartesianas de $\hat{x}$ (ver figura 1.31).

Para obtener unas coordenadas esféricas de un vector $\hat{x}$ a partir de sus coordenadas cartesianas (x, y, z), procedemos de la siguiente forma: en general, por la definición de la coordenada ρ, tenemos que

$$\rho = \sqrt{x^2 + y^2 + z^2}$$

identidad que también se deduce de las identidades anteriores.

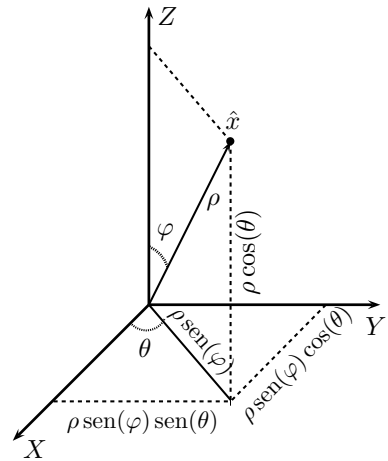


Figura 1.31: Recurriendo a las funciones trigonométricas, si (ρ, θ, φ) son coordenadas esféricas de un vector $\hat{x}$, deducimos que la terna $(x, y, z) = (\rho\,\mathrm{sen}(\varphi)\cos(\theta), \rho\,\mathrm{sen}(\varphi)\,\mathrm{sen}(\theta), \rho\cos(\varphi))$ son sus correspondientes coordenadas cartesianas.

Para determinar el ángulo φ basta con analizar dos casos; en el caso en que $x = 0 = y$, entonces hacemos $\varphi = 0$ si $z \geq 0$ y $\varphi = \pi$ si $z < 0$. Si $x^2 + y^2 > 0$, nuevamente por las ecuaciones 1.13 concluimos que se debe cumplir que

$$\begin{aligned}\frac{z^2}{x^2+y^2} &= \frac{(\rho\cos(\varphi))^2}{(\rho\,\mathrm{sen}(\varphi)\,\mathrm{sen}(\theta))^2 + (\rho\,\mathrm{sen}(\varphi)\cos(\theta))^2} \\ &= \frac{(\rho\cos(\varphi))^2}{(\rho\,\mathrm{sen}(\varphi))^2} \\ &= \tan^2(\varphi),\end{aligned}$$

de forma que

$$|\tan(\varphi)| = \frac{|z|}{\sqrt{x^2+y^2}}.$$

Si $z \geq 0$ bastará con tomar

$$\varphi = \arctan\left(\frac{z}{\sqrt{x^2+y^2}}\right).$$

Y si $z < 0$, bastará con tomar

$$\varphi = \frac{\pi}{2} - \arctan\left(\frac{z}{\sqrt{x^2+y^2}}\right),$$

en donde nuevamente arctan es la inversa de la función tangente que toma sus valores entre $-\pi/2$ y $\pi/2$.

En cuanto al ángulo θ, si $x = 0$ y $y \geq 0$, hacemos $\theta = \pi/2$, y si $y < 0$, entonces tomamos $\theta = 3\pi/2$. Cuando $x \neq 0$, por las ecuaciones 1.13 debemos tener que

$$\begin{aligned}\frac{y}{x} &= \frac{\cos(\theta)}{\operatorname{sen}(\theta)} \\ &= \tan(\theta)\end{aligned}$$

de modo que, en general, podemos obtener θ de la misma manera en que lo hicimos para el caso de las coordenadas polares.

1.7. Problemas

1. Pruebe las proposiciones 1.1 y 1.2.

2. Pruebe que las normas definidas en 1.4 son en efecto normas, es decir, que satisfacen las propiedades dadas en la proposición 1.1.

3. Pruebe que, si $\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, entonces $|x_i| \leq \|\hat{x}\|$, $|x_i| \leq \|\hat{x}\|_1$ y $|x_i| \leq \|\hat{x}\|_\infty$ para $i = 1, \ldots, n$.

4. Sean $a_1, \ldots, a_n \in \mathbb{R}$. Pruebe que $\left(\sum_{i=1}^{n} a_i\right)^2 \leq n \sum_{i=1}^{n} a_i^2$.

5. Pruebe que, si $\hat{x}_1, \ldots, \hat{x}_k \in \mathbb{R}^n$, entonces

$$\|\hat{x}_1 + \cdots + \hat{x}_k\| \leq \|\hat{x}_1\| + \cdots + \|\hat{x}_k\|$$

(recuerde que en el texto, la desigualdad del triángulo sólo se probó para dos vectores).

6. Sean $\hat{x}_1, \ldots, \hat{x}_k \in \mathbb{R}^n \setminus \{\hat{0}\}$ tales que $\hat{x}_i \cdot \hat{x}_j = 0$ si $i, j \in \{1, \ldots, k\}, i \neq j$. Pruebe que:

$$\|\hat{x}_1 + \cdots + \hat{x}_k\|^2 = \|\hat{x}_1\|^2 + \cdots + \|\hat{x}_k\|^2.$$

(Este resultado es conocido como el teorema de Pitágoras. Nótese que $k \leq n$. ¿Por qué?).

7. Sean $\hat{x}, \hat{y} \in \mathbb{R}^n$. Pruebe e interprete geométricamente los siguientes resultados:

   a) $\hat{x} \cdot \hat{y} = 0$ si y sólo si $\|\hat{x} + \hat{y}\| = \|\hat{x} - \hat{y}\|$
   b) $\hat{x} \cdot \hat{y} > 0$ si y sólo si $\|\hat{x} + \hat{y}\| > \|\hat{x} - \hat{y}\|$
   c) $\hat{x} \cdot \hat{y} < 0$ si y sólo si $\|\hat{x} + \hat{y}\| < \|\hat{x} - \hat{y}\|$

8. Sean $\hat{x}, \hat{y} \in \mathbb{R}^n$ diferentes de $\hat{0}$. Pruebe e interprete geométricamente los siguientes resultados:

   a) si $\|\hat{x}\| = \|\hat{y}\| = \|\hat{x} - \hat{y}\|$, entonces el ángulo entre $\hat{x}$ y $\hat{y}$ es $\pi/3$;
   b) si $\|\hat{x}\| = \|\hat{x} - \hat{y}\|$, entonces el ángulo entre $\hat{x}$ y $\hat{y}$ es igual al ángulo entre $\hat{y}$ y $\hat{y} - \hat{x}$.

9. Sean $\hat{x}, \hat{y} \in \mathbb{R}^n$. Pruebe que:

   a) $\|\hat{x} + \hat{y}\| = \|\hat{x}\| + \|\hat{y}\|$ si y sólo si existe $\lambda \in \mathbb{R}$, $\lambda > 0$, tal que $\hat{x} = \lambda \hat{y}$
   b) $\|\hat{x} - \hat{y}\| = \|\hat{x}\| + \|\hat{y}\|$ si y sólo si existe $\lambda \in \mathbb{R}$, $\lambda < 0$, tal que $\hat{x} = \lambda \hat{y}$
   c) $\|\hat{x} + \hat{y}\|^2 + \|\hat{x} - \hat{y}\|^2 = 2\left(\|\hat{x}\|^2 + \|\hat{y}\|^2\right)$

d) $|\|\hat{x}\| - \|\hat{y}\|| \leq \|\hat{x} \pm \hat{y}\|$.

10. Pruebe la proposición 1.5.

11. Sean $\hat{x}, \hat{y} \in \mathbb{R}^n$ y $r > 0$ tales que $\hat{y} \in B_r(\hat{x})$. Si $\hat{x} = (x_1, \ldots, x_n)$ y $\hat{y} = (y_1, \ldots, y_n)$, hacemos

$$\hat{x}_i = (y_1, \ldots, y_i, x_{i+1}, \ldots, x_n) \qquad \text{y} \qquad \hat{y}_i = (x_1, \ldots, x_i, y_{i+1}, \ldots, y_n)$$

para cada $i \in \{1, \ldots, n-1\}$, $\hat{x}_0 = \hat{y}_n = \hat{x}$ y $\hat{y}_0 = \hat{x}_n = \hat{y}$. Pruebe que:

a) $\hat{x}_i, \hat{y}_i \in B_r(\hat{x})$ para cada $i \in \{1, \ldots, n-1\}$

b) si $\hat{\xi}_i = \hat{x}_{i-1} + \xi(\hat{x}_i - \hat{x}_{i-1})$ y $\hat{\eta}_i = \hat{y}_{i-1} + \eta(\hat{y}_i - \hat{y}_{i-1})$ con $\xi, \eta \in (0,1)$, pruebe que

$$\left\|\hat{\xi}_i - \hat{x}\right\| \leq \|\hat{y} - \hat{x}\| \qquad \text{y} \qquad \|\hat{\eta}_i - \hat{x}\| \leq \|\hat{y} - \hat{x}\|$$

para cada $i \in \{1, \ldots, n\}$

c) dibuje los puntos $\hat{x}_i, \hat{y}_i$ para el caso de $\mathbb{R}^2$ y $\mathbb{R}^3$.

12. Sea $r > 0$ y $\hat{x} \in \mathbb{R}^n$. Definimos $B_r^{(1)}(\hat{x}) = \{\hat{y} \in \mathbb{R}^n \mid \|\hat{y} - \hat{x}\|_1 < r\}$ y $B_r^{(\infty)}(\hat{x}_0) = \{\hat{y} \in \mathbb{R}^n \mid \|\hat{y} - \hat{x}\|_\infty < r\}$.

a) Describa geométricamente los conjuntos $B_r^{(1)}(\hat{0})$ y $B_r^{(\infty)}(\hat{0})$ cuando $n = 2$ y $n = 3$.

b) En la definición 1.7, sustituya $B_r(\hat{x})$ por $B_r^{(1)}(\hat{x})$ y por $B_r^{(\infty)}(\hat{x})$ para definir $\text{int}_1(A)$, $\text{int}_\infty(A)$, $\text{ext}_1(A)$, $\text{ext}_\infty(A)$, $\text{Fr}_1(A)$ y $\text{Fr}_\infty(A)$, respectivamente.
Pruebe que $\text{int}_1(A) = \text{int}_\infty(A) = \text{int}(A)$, $\text{ext}_1(A) = \text{ext}_\infty(A) = \text{ext}(A)$, y $\text{Fr}_1(A) = \text{Fr}_\infty(A) = \text{Fr}(A)$.

13. Sean A, B subconjuntos de $\mathbb{R}^n$. Diga si cada una de las siguientes identidades y contenciones son ciertas. Pruebe su respuesta.

a) $\text{int}(A \cup B) = \text{int}(A) \cup \text{int}(B)$; $\text{int}(A \cap B) = \text{int}(A) \cap \text{int}(B)$

b) $\text{ext}(A \cup B) = \text{ext}(A) \cap \text{ext}(B)$; $\text{ext}(A \cap B) = \text{ext}(A) \cup \text{ext}(B)$

c) $\text{Fr}(A \cup B) = \text{Fr}(A) \cup \text{Fr}(B)$; $\text{Fr}(A \cap B) = \text{Fr}(A) \cap \text{Fr}(B)$

d) si $A \subset B$, entonces: $\text{int}(A) \subset \text{int}(B)$; $\text{ext}(B) \subset \text{ext}(A)$; $\text{Fr}(A) \subset \text{Fr}(B)$.

14. Sea $r > 0$ y $\hat{x} \in \mathbb{R}^n$. Si $A = B_r(\hat{x})$, pruebe que:

a) $\{\hat{y} \in \mathbb{R}^n \mid \|\hat{y} - \hat{x}\| > r\} \subset \text{ext}(A)$

b) $\{\hat{y} \in \mathbb{R}^n \mid \|\hat{y} - \hat{x}\| = r\} \subset \text{Fr}(A)$

c) las contenciones de los incisos anteriores son identidades.

15. Pruebe que, si $A \subset \mathbb{R}^n$ es un conjunto cerrado, entonces $\text{int}(\text{Fr}(A)) = \varnothing$.

16. Pruebe que el conjunto A del ejemplo 1.11 es un conjunto abierto.

17. Sean $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ tales que $a_i < b_i$ para $i = 1, \ldots, n$. Pruebe que el conjunto

$$\begin{aligned} A &= (a_1, b_1) \times \cdots \times (a_n, b_n) \\ &= \{(x_1, ..., x_n) \in \mathbb{R}^n \mid a_i < x_i < b_i, i = 1, ..., n\} \end{aligned}$$

es un conjunto abierto.

18. Sean $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ tales que $a_i \leq b_i$ para $i = 1, \ldots, n$. Pruebe que el conjunto

$$\begin{aligned} A &= [a_1, b_1] \times \cdots \times [a_n, b_n] \\ &= \{(x_1, ..., x_n) \in \mathbb{R}^n \mid a_i < x_i < b_i, i = 1, ..., n\} \end{aligned}$$

es un conjunto cerrado.

19. Sean $A \subset \mathbb{R}^n$ y $B \subset \mathbb{R}^m$. Si $A \times B = \{(\hat{x}, \hat{y}) \in \mathbb{R}^{n+m} \mid \hat{x} \in A, \hat{y} \in B\}$, pruebe que:

a) si A y B son abiertos (en $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente), entonces $A \times B$ es abierto (en $\mathbb{R}^{n+m}$)

b) si A y B son cerrados (en $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente), entonces $A \times B$ es cerrado (en $\mathbb{R}^{n+m}$)

c) si A y B son acotados (en $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente), entonces $A \times B$ es acotado (en $\mathbb{R}^{n+m}$).

20. Sea $U \subset \mathbb{R}^n$ un conjunto abierto no vacío y $\mathbb{Q}^n := \mathbb{Q} \times \cdots \times \mathbb{Q}$ ($\mathbb{Q}$ multiplicado n veces). Pruebe que:

    a) $U \cap \mathbb{Q}^n \neq \varnothing$

    b) U se puede poner como la unión de bolas (o vecindades) con centro en $\mathbb{Q}^n$ y radio racional.

21. Sean $A_1, \ldots, A_k$ subconjuntos de $\mathbb{R}^n$. Pruebe que:

    a) si cada A_i es abierto, entonces $A_1 \cup \cdots \cup A_k$ y $A_1 \cap \cdots \cap A_k$ son abiertos,

    b) si cada A_i es cerrado, entonces $A_1 \cup \cdots \cup A_k$ y $A_1 \cap \cdots \cap A_k$ son cerrados.

    ¿Estas afirmaciones siguen siendo ciertas para un número infinito de conjuntos? Pruebe su respuesta.

22. Pruebe la proposición 1.9 y muestre con un ejemplo que las contenciones que se dan ahí pueden ser propias.

23. Sean A, B subconjuntos de $\mathbb{R}^n$. Diga si las siguientes afirmaciones son ciertas. Pruebe su respuesta.

    a) Si $A \subset B$, entonces $A' \subset B'$

    b) $(A \cup B)' = A' \cup B'$

    c) $(A \cap B)' = A' \cap B'$.

24. Si $A = ([0,1] \times [0,1]) \cap (\mathbb{Q} \times \mathbb{Q}) = \{(x,y) \in \mathbb{R}^2 \mid x, y \in \mathbb{Q} \text{ y } 0 \leq x \leq 1, 0 \leq y \leq 1\}$.

    a) ¿Quién es A' y $\bar{A}$? Pruebe sus respuestas.

b) ¿A es abierto o cerrado? Pruebe sus respuestas.

25. Si $A = \{(m, 0) \in \mathbb{R}^2 \mid m \in \mathbb{Z}\}$

a) ¿quién es $\text{int}(A)$, $\text{Fr}(A)$, $\text{ext}(A)$, A' y $\bar{A}$? Pruebe sus respuestas.

b) ¿A es abierto o cerrado? Pruebe sus respuestas.

26. En la definición 1.10, sustituya $B_r(\hat{x})$ por $B_r^{(1)}(\hat{x})$ y por $B_r^{(\infty)}(\hat{x})$ para definir A_1' y A_∞', respectivamente. Pruebe que $A_1' = A_\infty' = A'$.

27. Sea A un subconjunto de $\mathbb{R}^n$. Pruebe que:

a) $\hat{x} \in \bar{A}$ si y sólo si para todo $r > 0$ se tiene que $B_r(\hat{x}) \cap A \neq \varnothing$

b) A es cerrado si y sólo si $A = \bar{A}$

c) A es cerrado si y sólo si $A' \subset A$

d) $A \cup A' = \bar{A}$

e) $\left(\bar{A}\right)^c = \text{ext}(A)$

f) $\text{int}(\text{int}(A)) = \text{int}(A)$ y $\overline{(\bar{A})} = \bar{A}$

g) $\text{Fr}(A) = \bar{A} \cap \overline{(A^c)}$.

28. Sea A un subconjunto de $\mathbb{R}^n$. Pruebe que:

a) si $B \subset A$ y B es abierto, entonces $B \subset \text{int}(A)$ (es decir, de los conjuntos abiertos que están contenidos en A, $\text{int}(A)$ es el más "grande")

b) si $A \subset B$ y B es cerrado, entonces $\bar{A} \subset B$ (es decir, de los conjuntos cerrados que contienen a A, $\bar{A}$ es el más "chico").

29. Sean A, B subconjuntos de $\mathbb{R}^n$. Diga si las siguientes afirmaciones son ciertas. Pruebe sus respuestas.

a) si $A \subset B$, entonces $\bar{A} \subset \bar{B}$

b) $\overline{(A \cup B)} = \bar{A} \cup \bar{B}$

c) $\overline{(A \cap B)} \subset \bar{A} \cap \bar{B}$

d) $\overline{(A \cap B)} = \bar{A} \cap \bar{B}$.

30. Sea $A \subset \mathbb{R}^n$. Pruebe que: A es un conjunto acotado si y sólo si para todo $\hat{x} \in \mathbb{R}^n$ existe $M > 0$ (que depende de $\hat{x}$) tal que $\|\hat{x} - \hat{y}\| \leq M$ para todo $\hat{y} \in A$.

31. Pruebe que el conjunto A definido en cada uno de los problemas 17 y 18 es un conjunto acotado.

32. Sea $A \subset \mathbb{R}$, $A \neq \varnothing$. Pruebe que si A es cerrado y acotado, entonces $\text{ínf}(A), \sup(A) \in A$.

33. Sea $A \subset \mathbb{R}^n$ un conjunto infinito. Pruebe que si A es un conjunto cerrado y acotado, entonces todo subconjunto infinito B de A tiene un punto de acumulación en A.

34. Sea $\{A_k\}$ una sucesión de subconjuntos de $\mathbb{R}^n$, cerrados, acotados y no vacíos, tales que $A_{k+1} \subset A_k$ para toda $k \in \mathbb{N}$ (es decir, lo que se llama una sucesión "anidada" de conjuntos). Pruebe que

$$\bigcap_{k \in \mathbb{N}} A_k \neq \varnothing.$$

Este resultado sigue siendo cierto si los conjuntos ¿no son cerrados? o ¿no son acotados? Pruebe sus respuestas.

35. Sea $A \subset \mathbb{R}^n$ tal que $A' \neq \varnothing$. Pruebe que para todo $\varepsilon > 0$ existen $\hat{x}, \hat{y} \in A$ tales que $0 < \|\hat{x} - \hat{y}\| < \varepsilon$.

36. Determine si la siguiente proposición es verdadera: *si* $A \subset \mathbb{R}^n$ *es un conjunto infinito y para todo* $\varepsilon > 0$ *existen* $\hat{x}, \hat{y} \in A$ *tales que* $0 < \|\hat{x} - \hat{y}\| < \varepsilon$, *entonces* $A' \neq \varnothing$. Pruebe su respuesta.

37. Sean $A \subset \mathbb{R}^n$ un conjunto infinito y $c \in \mathbb{R}$, con $c > 0$. Pruebe que, si $\|\hat{x} - \hat{y}\| \geq c$ para todo $\hat{x}, \hat{y} \in A$, entonces A es un conjunto no acotado.

38. Sea $A \subset \mathbb{R}^n$ tal que $A' = \varnothing$. Pruebe que si $M > 0$ y $A_M = \{\hat{x} \in A \mid \|\hat{x}\| \leq M\}$, entonces A_M es finito.

39. Sean $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ tales que $a_i < b_i$ para $i = 1, \ldots, n$. Pruebe que el conjunto

$$\begin{aligned} A &= (a_1, b_1) \times \cdots \times (a_n, b_n) \\ &= \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i < x_i < b_i, i = 1, \ldots, n\} \end{aligned}$$

es un conjunto convexo.

40. Sea $A \subset \mathbb{R}$, $A \neq \varnothing$. Pruebe que las siguientes afirmaciones son equivalentes:

    a) A es conexo.

    b) Si para todos $a, b \in A$, $a < b$, y $c \in \mathbb{R}$ tal que $a < c < b$, entonces $c \in A$.

    c) A es un intervalo (es decir, A es de alguna de las siguientes formas: $(-\infty, b)$, $(-\infty, b]$, (a, b), $(a, b]$, $[a, b)$,$[a, b]$, (a, ∞), $[a, \infty)$ o $(-\infty, \infty) = \mathbb{R}$).

41. Pruebe la proposición 1.11.

42. Sea $\varnothing \neq A \subsetneq \mathbb{R}^n$. Pruebe que:

    a) A no puede ser abierto y cerrado a la vez

    b) $\operatorname{Fr}(A) \neq \varnothing$

    c) si $\hat{x} \in A$ y $\hat{y} \in A^c$, entonces $[\hat{x}, \hat{y}] \cap \operatorname{Fr}(A) \neq \varnothing$.

43. Sea $A \subset \mathbb{R}^n$ abierto, $A \neq \varnothing$. Pruebe que A es disconexo sí y sólo si existen $B, C \subset \mathbb{R}^n$ tales que B y C son abiertos, ajenos, no vacíos y $A = B \cup C$.

44. Sea $A \subset \mathbb{R}^n$ cerrado, $A \neq \varnothing$. Pruebe que A es disconexo si y sólo si existen $B, C \subset \mathbb{R}^n$ tales que B y C son cerrados, ajenos, no vacíos y $A = B \cup C$.

45. Sean $A, B, C \subset \mathbb{R}^n$. Pruebe que si B y C están separados, entonces $C \cap A$ y $B \cap A$ están separados.

46. Sean $A, B, C \subset \mathbb{R}^n$ tales que $\varnothing \neq A \subset B \cup C$. Pruebe que, si A es conexo y B y C están separados, entonces $A \cap B = \varnothing$ o $A \cap C = \varnothing$ (y por lo tanto $A \subset C$ o $A \subset B$, respectivamente).

47. Sean $A, D \subset \mathbb{R}^n$ conjuntos conexos tales que $A \cap D \neq \varnothing$. Determine si las siguientes afirmaciones son ciertas. Pruebe su respuesta.

    a) $A \cap D$ es conexo.

    b) $A \cup D$ es conexo.

48. Sean $A, B \subset \mathbb{R}^n$ tales que $A \cap B$ y $A \cup B$ son conjuntos no vacíos y conexos. Determine si las siguientes afirmaciones son ciertas. Pruebe su respuesta.

    a) los conjuntos A y B son conexos,

    b) si A y B son conjuntos cerrados, entonces A y B son conexos,

    c) si A y B son conjuntos abiertos, entonces A y B son conexos.

49. ¿La proposición 1.13 es cierta si A no es abierto? Pruebe su respuesta.

50. Sea $A \subset \mathbb{R}^n$. A es un *conjunto estrellado* si existe $\hat{x}_0 \in A$ tal que para todo $\hat{x} \in A$ se satisface que $[\hat{x}_0, \hat{x}] \subset A$ (en cuyo caso se dice que A es estrellado con respecto de $\hat{x}_0$). Pruebe que todo conjunto estrellado es conexo.

51. Sean $\hat{x}_1$ y $\hat{x}_2$ vectores en el plano, y (ρ_1, θ_1) y (ρ_2, θ_2) coordenadas polares de éstos, respectivamente. Si $\hat{x}_1 + \hat{x}_2$ y $\lambda \hat{x}_1$ ($\lambda \in \mathbb{R}$) son la suma y producto por un escalar que se definieron geométricamente en el texto, encuentre coordenadas polares para estos vectores en términos de las coordenadas polares (ρ_1, θ_1) y (ρ_2, θ_2). Compruebe su respuesta convirtiendo a las coordenadas cartesianas correspondientes.

Capítulo 2

Funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$

Al inicio del capítulo anterior dimos una lista de ejemplos que planteaban diversas situaciones. Nuestro objetivo principal fue mostrar, para cada una de estas situaciones, que la función que permitía describirla era una función cuyas variables independiente o dependiente (o ambas) pertenecían a algún espacio vectorial, y por esta razón dichas variables siempre eran suceptibles de representarse (describirse o medirse) por una cierta cantidad de números reales (dos, tres, cuatro ¡o más!), es decir, por una n-ada de números reales.

De esta forma, cada función que surge de estos ejemplos se puede considerar, en última instancia, como una función cuya variable independiente pertenece a algún subconjunto de $\mathbb{R}^n$ y la variable dependiente a algún subconjunto de $\mathbb{R}^m$. Por lo anterior, las m-adas que representen a la variable dependiente estarán expresadas en términos de las n-adas que representen a la variable independiente.

En el capítulo anterior nos concentramos en estudiar al conjunto de las n-adas, es decir $\mathbb{R}^n$. Por las razones anteriormente expuestas, en este capítulo realizaremos un análisis más detallado de las funciones definidas sobre subconjuntos de $\mathbb{R}^n$ cuyos valores estén en $\mathbb{R}^m$, sin importar (por ahora) a qué espacios vectoriales pertenezcan las variables (independientes y dependientes) representadas por estas n-adas (y m-adas) de números reales.

2.1. Álgebra y geometría de las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$

Iniciamos esta sección estableciendo la nomenclatura y la notación con la cual trabajeremos a lo largo de todo este texto. Casi siempre usaremos la letra f para denotar a una función. Si está definida sobre un conjunto $A \subset \mathbb{R}^n$ y toma valores en $\mathbb{R}^m$, todo ello lo escribiremos de la siguiente forma:

$$f : A \subset \mathbb{R}^n \to \mathbb{R}^m.$$

Que los valores de f estén en $\mathbb{R}^m$, significa que para cada $\hat{x} \in A$ se debe tener que $f(\hat{x}) \in \mathbb{R}^m$, razón por la que el valor $f(\hat{x})$ debe de tener m coordenadas, a las cuales denotaremos por $f_i(\hat{x}) \in \mathbb{R}$, con $i = 1, \ldots, m$. Es decir, $f(\hat{x}) = (f_1(\hat{x}), \ldots, f_m(\hat{x}))$. De esta forma, toda función $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ determina (o está determinada por) m funciones $f_i : A \subset \mathbb{R}^n \to \mathbb{R}$, a las cuales conoceremos con el nombre de *funciones coordenadas*. Con frecuencia escribiremos que $f = (f_1, \ldots, f_m)$.

Las funciones coordenadas siempre son funciones de valores reales y veremos que tienen un papel muy importante en los conceptos y resultados que desarrollaremos a lo largo de este texto.

La siguiente lista ejemplifica el tipo de funciones con las cuales trabajaremos.

Ejemplo 2.1 *Considere las siguientes funciones:*

1. $f : \mathbb{R}^n \to \mathbb{R}$ *definida como* $f(\hat{x}) = \|\hat{x}\|$
2. $f : [0, 2\pi] \subset \mathbb{R} \to \mathbb{R}^2$ *definida como* $f(t) = (\cos(t), \operatorname{sen}(t))$
3. $f : A = [0, 2\pi] \times [0, \pi] \subset \mathbb{R}^2 \to \mathbb{R}^3$ *definida como*
$$f(x, y) = (\cos(x)\operatorname{sen}(y), \operatorname{sen}(x)\operatorname{sen}(y), \cos(y))$$
4. $f : A = \mathbb{R}^2 \setminus \{(0, 0)\} \subset \mathbb{R}^2 \to \mathbb{R}^2$ *definida como*
$$f(x, y) = \left(\frac{x}{\sqrt{x^2 + y^2}}, \frac{y}{\sqrt{x^2 + y^2}} \right)$$

5. $f : A = \mathbb{R}^2 \setminus \{(x, y) \in \mathbb{R}^2 \mid x \leq 0 \text{ y } y = 0\} \subset \mathbb{R}^2 \to \mathbb{R}$ *definida como*

$$f(x,y) = \begin{cases} 0 & \text{si } 0 \leq x \\ x^2 & \text{si } x \leq 0 \text{ y } 0 < y \\ -x^2 & \text{si } x \leq 0 \text{ y } y < 0 \end{cases}$$

Otro aspecto que mencionaremos rápidamente, es el relacionado con las operaciones algebraicas que se pueden realizar con este tipo de funciones y que serán con las que trabajaremos en este texto. Todas estas operaciones las formalizamos en la siguiente

Definición 2.1 *Sean $f, g : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $c \in \mathbb{R}$, $\hat{c} \in \mathbb{R}^m$ y $h : D \subset \mathbb{R}^m \to \mathbb{R}^k$. Definimos:*

1. *La suma de f y g, que denotamos por $f + g$, como $(f + g)(\hat{x}) := f(\hat{x}) + g(\hat{x})$ para cada $\hat{x} \in A$.*

2. *El producto del escalar c por la función f, que denotamos por cf, como $(cf)(\hat{x}) := cf(\hat{x})$ para cada $\hat{x} \in A$.*

3. *El producto punto de $\hat{c}$ por f, que denotamos por $\hat{c}\cdot f$, como $(\hat{c}\cdot f)(\hat{x}) := \hat{c} \cdot f(\hat{x})$ para cada $\hat{x} \in A$.*

4. *En general, el producto punto de f por g, que denotamos por $f \cdot g$, como $(f \cdot g)(\hat{x}) := f(\hat{x}) \cdot g(\hat{x})$ para cada $\hat{x} \in A$.*

5. *Si $m = 3$, el producto cruz de f por g, que denotamos por $f \times g$, como $(f \times g)(\hat{x}) := f(\hat{x}) \times g(\hat{x})$ para cada $\hat{x} \in A$.*

6. *Si $m = 1$, el cociente de f entre g, que denotamos por f/g, como $(f/g)(\hat{x}) := f(\hat{x})/g(\hat{x})$ para cada $\hat{x} \in B = \{\hat{x} \in A \mid g(\hat{x}) \neq 0\}$.*

7. *La composición de h con f, que denotamos por $h\circ f$, como $(h\circ f)(\hat{x}) := h(f(\hat{x}))$ para cada $\hat{x} \in B = \{\hat{x} \in A \mid f(\hat{x}) \in D\}$.*

Obtener algún tipo de representación geométrica de una función de $\mathbb{R}^n$ en $\mathbb{R}^n$, depende del tamao de n y m. En algunos casos, sobre todo cuando n y m no son muy grandes, existe la posibilidad de "hacer" un poco de geometría con este tipo de funciones. Para lograr lo anterior, asociados a una función $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, definiremos algunos conjuntos a los cuales haremos alusión a lo largo de todo este texto. El primero de ellos, sin importar qué tan grandes sean n y m, será lo que llamaremos *la gráfica de* f.

Definición 2.2 *Dada $f = (f_1, \ldots, f_m) : A \subset \mathbb{R}^n \to \mathbb{R}^m$ definimos la gráfica de f, que denotaremos por G_f, como el siguiente conjunto:*

$$G_f := \{(x_1, \ldots, x_n, f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)) \in \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m} \mid (x_1, \ldots, x_n) \in A\}.$$

y que con frecuencia escribiremos simplemente (para evitar expresiones muy largas, pero no sin cierto abuso de notación) como

$$G_f := \{(\hat{x}, f(\hat{x})) \in \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m} \mid \hat{x} \in A\}$$

Seguramente el lector estará de acuerdo en que sólo podremos "dibujar" (¡y ver!) la gráfica de una función si $2 \leq n + m \leq 3$. Es decir, en muy pocos casos. Peor aún, en el caso $n = 1$ y $m = 2$ este conjunto no resulta de mucha utilidad, y como el caso $n = 1$ y $m = 1$ ya se estudió con mucho cuidado en los cursos previos de cálculo de una variable, sólo en el caso $n = 2$ y $m = 1$ podremos dibujar la gráfica, lo que no le quita interés a su estudio.

De acuerdo con lo anterior, el tipo de función para el cual vale la pena destacar algunas estrategias que nos permitan darnos una idea geométrica de cómo es su gráfica (independientemente de que hoy en día se cuenta con herramientas muy sofisticadas para dibujarlas), será cuando tengamos una función $f : A \subset \mathbb{R}^2 \to \mathbb{R}$. En este caso, la gráfica de f se puede escribir como

$$G_f = \{(x, y, z) \in \mathbb{R}^3 \mid z = f(x, y) \text{ y } (x, y) \in A\}$$

lo que en principio establece que los elementos de la gráfica deben satisfacer la ecuación

$$z = f(x, y),$$

con la restricción de que $(x, y) \in A$. De esta forma, la experiencia del lector (obtenida en sus cursos de geometría analítica) para visualizar conjuntos en $\mathbb{R}^3$ definidos a partir de una ecuación (la mayoría de las veces de tipo cuadrático en las variables x, y y z), le será de gran utilidad.

Relacionado con lo anterior, de la misma forma que no cualquier subconjunto de $\mathbb{R}^2$ puede ser la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$, no cualquier subconjunto de $\mathbb{R}^3$ puede ser la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$.

Seguramente el lector recordará que para una función definida de $\mathbb{R}$ (o de un subconjunto de $\mathbb{R}$) en $\mathbb{R}$, la gráfica de f es un subconjunto de $\mathbb{R}^2$ que intersecta a cualquier recta paralela al eje Y en a lo más un punto. En el caso de funciones de $\mathbb{R}^2$ en $\mathbb{R}$, hay un "criterio geométrico" equivalente: la gráfica de una función definida de $\mathbb{R}^2$ (o de un subconjunto de $\mathbb{R}^2$) en $\mathbb{R}$ es un subconjunto de $\mathbb{R}^3$ que intersecta a cualquier recta paralela al eje Z en a lo más un punto.

Como el lector recordará también, la justificción de este criterio se basa en el hecho de que estamos hablando de funciones, de tal forma que a cada elemento del dominio le corresponde uno y sólo un valor del contradominio.

En el siguiente ejemplo ilustramos cómo usar este criterio geométrico, para identificar o esbozar la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$.

Ejemplo 2.2 *Esboce la gráfica de la función* $f : A \subset \mathbb{R}^2 \to \mathbb{R}$, *con* $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$, *definida como*

$$f(x, y) = \sqrt{1 - (x^2 + y^2)}.$$

Con base en la observación que hicimos anteriormente, las ternas $(x, y, z) \in G_f$ *deben ser tales que*

$$\begin{aligned} z &= f(x, y) \\ &= \sqrt{1 - (x^2 + y^2)}, \end{aligned} \tag{2.1}$$

de tal forma que, tomando el cuadrado en ambos lados de esta identidad (operación que será muy importante recordar más adelante), concluimos que estas ternas satisfacen la ecuación

$$z^2 = 1 - (x^2 + y^2)$$

o equivalentemente, la ecuación

$$x^2 + y^2 + z^2 = 1, \tag{2.2}$$

que seguramente el lector reconocerá como la ecuación de una esfera de radio 1 *con centro en el origen* $(0, 0, 0)$. *Pero justo por la observación anterior a este ejemplo, dado que muchas rectas paralelas al eje Z (incluyendo el propio eje) intersectan a esta esfera en más de un punto, la gráfica de nuestra función no puede ser toda la esfera.*

Para determinar qué parte de la esfera corresponde a la gráfica de nuestra función, recordemos que la ternas que están en la gráfica satisfacen la ecuación 2.1, de donde se deduce que la coordenada z de estas ternas siempre es mayor o igual a 0 *(a diferencia de las ternas que satisfacen la ecuación 2.2, de la que no se puede deducir que la coordenada z tenga que cumplir con la misma condición). Como seguramente el lector ya lo habrá notado, esta "pérdida" del signo de la coordenada z fue producto de haber tomado el cuadrado en la ecuación 2.1.*

Por otra parte, si $(x, y, z) \in \mathbb{R}^3$, *con* $z \geq 0$, *satisface la ecuación 2.2, entonces se tiene que*

$$z = \sqrt{1 - (x^2 + y^2)} \tag{2.3}$$

$$= f(x, y), \tag{2.4}$$

de donde se concluye que $(x, y, z) \in G_f$.

Resumiendo lo anterior, concluimos que la gráfica de nuestra función es la parte de la esfera unitaria que se encuentra por arriba (y sobre) el plano XY, como se muestra en la figura 2.1.

Con frecuencia haremos referencia a otro conjunto asociado a una función de $\mathbb{R}^n$ en $\mathbb{R}$, el llamado conjunto de nivel. Su definición formal es la siguiente:

Definición 2.3 *Sean* $f : A \subset \mathbb{R}^n \to \mathbb{R}$ *y* $c \in \mathbb{R}$. *Definimos el conjunto de nivel c de f, que denotamos por* $N_c(f)$, *como*

$$N_c(f) := \{\hat{x} \in A \mid f(\hat{x}) = c\}.$$

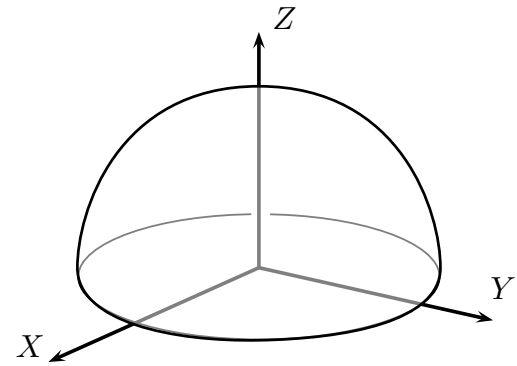


Figura 2.1: Gráfica de la función $f(x,y) = \sqrt{1-(x^2+y^2)}$.

Una propiedad evidente (pero importante) es que los conjuntos de nivel correspondientes a valores distintos de c, no se intersectan. Esta propiedad es una consecuencia inmediata del hecho de que f sea una función. La dejaremos plasmada en la siguiente

Observación 2.1 *Dados $c, d \in \mathbb{R}$ tales que $N_c(f) \neq \varnothing$ y $N_d(f) \neq \varnothing$, se tiene que*

$$N_c(f) \cap N_d(f) = \varnothing \text{ si y sólo si } c \neq d.$$

Como es de esperarse, los conjuntos de nivel sólo se pueden "ver" o "dibujar" si el dominio de nuestra función está contenido en $\mathbb{R}^2$ o $\mathbb{R}^3$ (lo que no reduce su importancia para dimensiones mayores). Por esta razón, daremos algunos ejemplos sólo en estos casos.

Ejemplo 2.3 *Determine los conjuntos de nivel de las siguientes funciones:*

1. *$f(x,y) = \sqrt{x^2+y^2}$ con $(x,y) \in \mathbb{R}^2$. Dado que $f(x,y) \geq 0$ para toda $(x,y) \in \mathbb{R}^2$, para $c < 0$ se tiene que $N_c(f) = \varnothing$. Si $c = 0$, entonces $N_c(f) = \{(0,0)\}$, y si $c > 0$, entonces*

$$N_c(f) = \left\{(x,y) \in \mathbb{R}^2 \mid \sqrt{x^2+y^2} = f(x,y) = c\right\},$$

que no es más que la circunferencia de radio c con centro en el origen (ver figura 2.2).

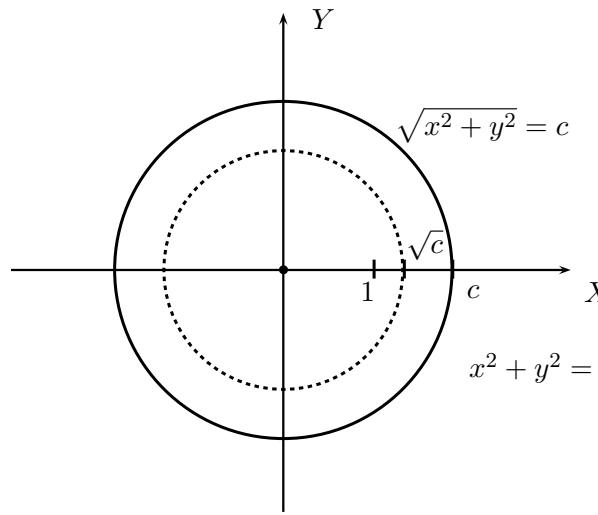


Figura 2.2: Curvas de nivel c de las funciones $f(x,y)=\sqrt{x^2+y^2}$ y $f(x,y)=x^2+y^2$, las circunferencias de radio c y $\sqrt{c}$, respectivamente.

2. $f(x,y)=x^2+y^2$ *con* $(x,y)\in\mathbb{R}^2$. *Como para esta función también se tiene que* $f(x,y)\geq 0$ *para toda* $(x,y)\in\mathbb{R}^2$, *si* $c<0$, *entonces* $N_c(f)=\varnothing$. *Asimismo, si* $c=0$, *entonces* $N_c(f)=\{(0,0)\}$, *y si* $c>0$, *entonces*

$$N_c(f)=\left\{(x,y)\in\mathbb{R}^2 \mid x^2+y^2=f(x,y)=c\right\},$$

que en este caso es la circunferencia de radio $\sqrt{c}$ *con centro en el origen (ver figura 2.2).*

3. $f(x,y,z)=x+y+z$ *con* $(x,y,z)\in\mathbb{R}^3$. *En este caso se cumple que* $N_c(f)\neq\varnothing$ *para toda* $c\in\mathbb{R}$ *y*

$$N_c(f)=\left\{(x,y,z)\in\mathbb{R}^3 \mid x+y+z=f(x,y)=c\right\}$$

que, como el lector reconocerá fácilmente, se trata de un plano con vector normal $(1, 1, 1)$ *(ver figura 2.3).*

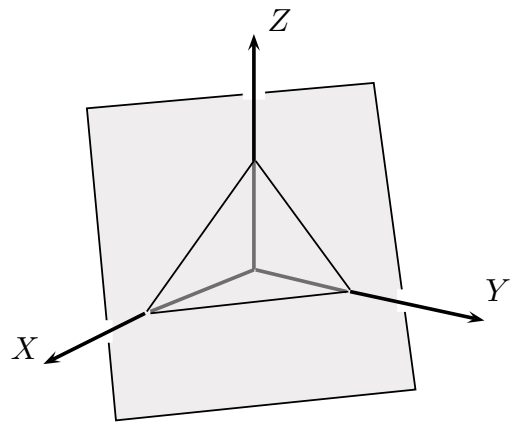


Figura 2.3: Conjunto de nivel 1 de la función $f(x, y, z) = x + y + z$.

Entre otras cosas, y para el caso de $\mathbb{R}^2$, este tipo de conjuntos también serán útiles para esbozar la gráfica de una función. En efecto, si cada conjunto de nivel $N_c(f)$ (o una cantidad suficiente de ellos) se traslada paralelamente al plano XY a la "altura" c, obtenemos un bosquejo de la gráfica de f (ver figura 2.4). Este hecho, junto con todas las demás técnicas que el lector haya aprendido en sus cursos de Geometría Analítica (como por ejemplo, la intersección con los "planos coordenados"), serán útiles a la hora de intentar visualizar la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$.

En realidad, los conjuntos de nivel son un caso particular de otros conjuntos con los cuales vamos a trabajar en este mismo capítulo, un poco más adelante. Aprovechamos esta relación para definir lo que se conoce como "*la imagen inversa de D bajo f*", de la siguiente manera:

Definición 2.4 *Sean* $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ *y* $D \subset \mathbb{R}^m$*. Definimos la imagen inversa de D bajo f, que denotamos por* $f^{-1}(D)$*, como el conjunto dado por:*

$$f^{-1}(D) := \{\hat{x} \in A \mid f(\hat{x}) \in D\}\,.$$

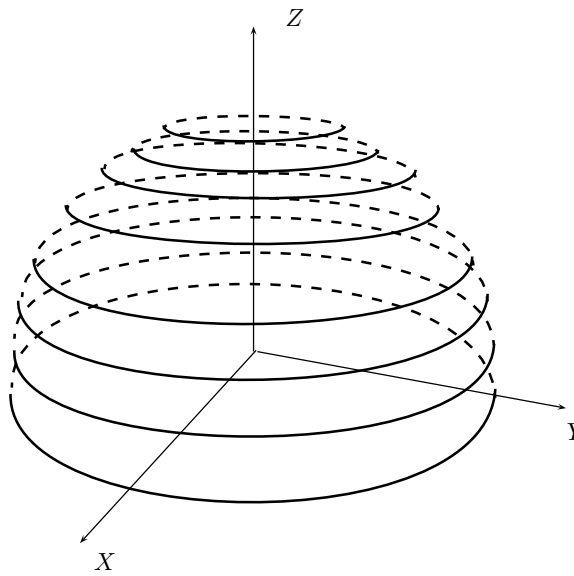


Figura 2.4: Bosquejo de la gráfica de la función $f(x,y) = \sqrt{1-(x^2+y^2)}$ a partir de algunas de sus curvas de nivel.

Esperemos que la notación usada en esta definición no cause confusión con la notación de función inversa; si lo que está entre paréntesis (el "argumento") es un conjunto, nos referimos a la imagen inversa; y si es un elemento de $\mathbb{R}^n$, hablamos de la función inversa de f.

Observación 2.2 *Como el lector podrá verificar fácilmente, si* $f : A \subset \mathbb{R}^n \to \mathbb{R}$ *y* $c \in \mathbb{R}$*, entonces*

$$N_c(f) = f^{-1}(\{c\}).$$

Para terminar la lista de conjuntos asociados con una función $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, dado un conjunto $B \subset A$, definiremos ahora lo que se conoce como "*la imagen directa* (o simplemente "*la imagen*") *de* B *bajo* f" de la siguiente manera.

Definición 2.5 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $B \subset A$. Definimos la imagen (directa) de B bajo f, que denotamos*[1] *por $f(B)$, como el conjunto dado por:*

$$f(B) := \{f(\hat{x}) \in \mathbb{R}^m \mid \hat{x} \in B\} = \{\hat{y} \in \mathbb{R}^m \mid \textit{existe } \hat{x} \in B \textit{ tal que } f(\hat{x}) = \hat{y}\}.$$

Desde un punto de vista geométrico, la imagen de un conjunto bajo una función $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ jugará un papel relevante cuando $n \leq m$. En estos casos, dado un conjunto $B \subset A$, será muy importante reconocer quién es el conjunto $f(B)$.

De hecho, apoyados en el concepto de imagen directa, diremos (por ahora de manera informal) que un subconjunto $C \subset \mathbb{R}^m$ que pueda obtenerse (todo él o "en partes") como la imagen de un conjunto bajo una función $f : I \subset \mathbb{R} \to \mathbb{R}^m$ (que cumpla con ciertas propiedades de derivabilidad que definiremos más adelante) es una *curva en* $\mathbb{R}^m$.

Análogamente, si un subconjunto $S \subset \mathbb{R}^m$ se puede obtener (todo él o "en partes") como la imagen de un conjunto bajo una función $f : A \subset \mathbb{R}^2 \to \mathbb{R}^m$ (que también cumpla con algunas propiedades de derivabilidad), nos referiremos a él como una *superficie en* $\mathbb{R}^m$.

Como es de suponer, los únicos casos en los que será posible dibujar (o bosquejar) la imagen de un conjunto bajo una función serán aquellos en los que el contradominio de la función es $\mathbb{R}^2$ o $\mathbb{R}^3$. Es decir, sólo para funciones de $\mathbb{R}$ en $\mathbb{R}^2$, de $\mathbb{R}$ en $\mathbb{R}^3$, de $\mathbb{R}^2$ en $\mathbb{R}^2$, de $\mathbb{R}^2$ en $\mathbb{R}^3$ y de $\mathbb{R}^3$ en $\mathbb{R}^3$.

En el capítulo 3 vamos a estudiar con más detalle las funciones de $\mathbb{R}$ en $\mathbb{R}^n$ (en particular en $\mathbb{R}^2$ y en $\mathbb{R}^3$), razón por la cual ahora sólo nos concentraremos en dar ejemplos de cómo es la imagen de algunos conjuntos bajo este tipo de funciones.

Ejemplo 2.4 *Sea $f : \mathbb{R} \to \mathbb{R}^2$ dada por $f(t) = (\cos(t), \operatorname{sen}(t))$. Describiremos (o bosquejaremos) los conjuntos $f([-\pi/2, \pi/2])$, $f([t_0, t_0 + 2\pi])$ y $f(\{t_0 + k\pi/2 \mid k \in \mathbb{Z}\})$ con $t_0 \in \mathbb{R}$ fijo, en ambos casos.*

[1] Como en el caso de la imagen inversa, nuevamente esperemos que esta notación no cause confusión con la notación de función; si lo que está entre paréntesis es un conjunto, nos referimos a la imagen directa; y si es un elemento de $\mathbb{R}^n$, hablamos de la función evaluada en ese elemento.

Para lograr nuestro objetivo, es importante observar lo siguiente: si interpretamos a la variable t como el ángulo dirigido (medido en radianes) formado por la parte positiva del eje X y una semirrecta que parte del origen, entonces el punto representado por la pareja $(\cos(t), \operatorname{sen}(t)) = f(t)$ es la intersección de esta semirrecta con la circunferencia de radio 1 con centro en el origen.

Tomando en consideración este hecho, ahora fácilmente podemos concluir que: $f([-\pi/2, \pi/2])$ es el conjunto formado por la parte de la circunferencia unitaria con centro en el origen que se encuentra en el semiplano derecho (figura 2.5 (a)), $f([t_0, t_0 + 2\pi))$ es dicha circunferencia completa (figura 2.5 (b)), y $f(\{t_0 + k\pi/2 \mid k \in \mathbb{Z}\})$ es un conjunto formado sólo por cuatro puntos (figura 2.5 (c)).

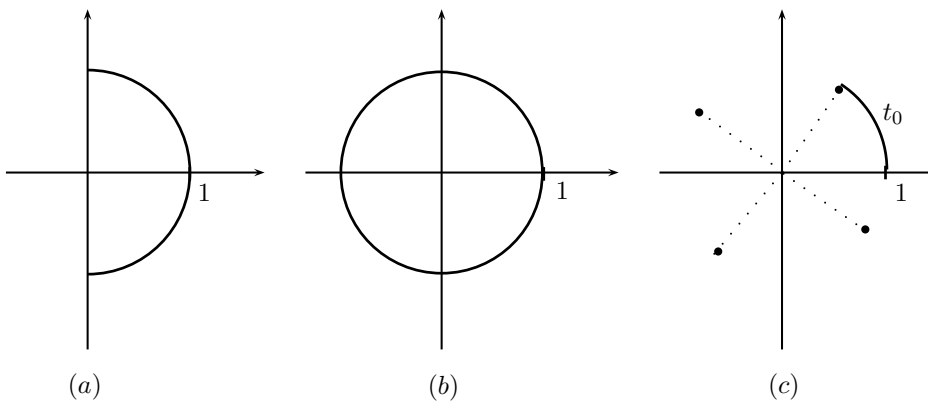


Figura 2.5: Los conjuntos $f([-\pi/2, \pi/2])$, $f([t_0, t_0 + 2\pi])$ y $f(\{t_0 + k\pi/2 \mid k \in \mathbb{Z}\})$.

Ejemplo 2.5 *Sea $f : A \subset \mathbb{R}^2 \to \mathbb{R}^3$, con $A = [0, 2\pi] \times [0, \pi]$, dada por*

$$f(t, s) = (\cos(t) \operatorname{sen}(s), \operatorname{sen}(t) \operatorname{sen}(s), \cos(s)).$$

Describiremos (o bosquejaremos) los conjuntos $f(\{(t, s_0) \mid t \in [0, 2\pi]\})$, $f(\{(t_0, s) \mid s \in [0, \pi]\})$ con $s_0 \in [0, \pi]$ y $t_0 \in [0, 2\pi]$ fijos.

De forma análoga al ejemplo anterior, ahora es importante notar que las ternas

$$(\cos(t)\,\mathrm{sen}(s), \mathrm{sen}(t)\,\mathrm{sen}(s), \cos(s))$$

satisfacen la ecuación $x^2 + y^2 + z^2 = 1$ *por lo que podemos concluir que los conjuntos que buscamos están contenidos en la esfera de radio* 1 *con centro en el origen.*

Como todas las ternas $f(t, s_0) = (\cos(t)\,\mathrm{sen}(s_0), \mathrm{sen}(t)\,\mathrm{sen}(s_0), \cos(s_0))$ *tienen la particularidad de que su tercera coordenada es la misma, entonces satisfacen la ecuación del plano* $z = \cos(s_0)$*, el cual es paralelo al plano* XY*. En virtud de la primera observación, concluimos que nuestro conjunto está contenido en la intersección de la esfera que mencionamos y este plano, que es una circunferencia de radio* $\mathrm{sen}(s_0)$ *con centro en el punto* $(0, 0, \cos(s_0))$*. Dado que* $t \in [0, 2\pi]$*, tenemos que el conjunto buscado coincide con esta circunferencia (ver figura 2.6).*

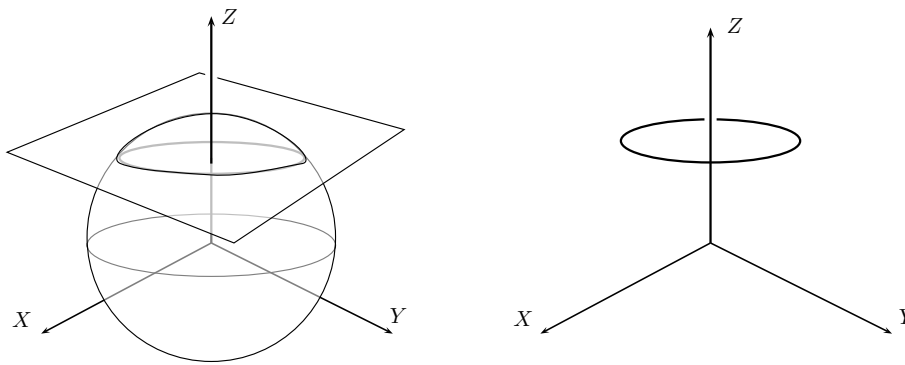


Figura 2.6: El conjunto $f(\{(t, s_0) \mid t \in [0, 2\pi]\})$ se obtiene al intersectar la esfera $x^2 + y^2 + z^2 = 1$ con el plano $z = \cos(s_0)$.

Por otra parte, como $\mathrm{sen}(s) \geq 0$ *para toda* $s \in [0, \pi]$*, se tiene que las parejas* $(\cos(t_0)\,\mathrm{sen}(s), \mathrm{sen}(t_0)\,\mathrm{sen}(s))$ *(que son la proyección sobre el plano* XY *de las ternas* $(\cos(t_0)\,\mathrm{sen}(s), \mathrm{sen}(t_0)\,\mathrm{sen}(s), \cos(s)) = f(t_0, s)$*) pertenecen a la semirrecta que parte del origen y que forma un ángulo de* t_0 *radianes con la parte positiva del eje* X*.*

Como las ternas $(\cos(t_0)\,\mathrm{sen}(s), \mathrm{sen}(t_0)\,\mathrm{sen}(s), \cos(s)) = f(t_0, s)$ *también satisfacen la ecuación del plano* $-\mathrm{sen}(t_0)x + \cos(t_0)y = 0$*, concluimos que el conjunto* $f(\{(t_0, s) \mid s \in [0, \pi]\})$ *está contenido en la intersección de la esfera unitaria con centro en el origen y este plano. Esta intersección es una circunferencia completa, pero dado que la proyección de los elementos de nuestro conjunto sólo caen en la semirecta antes descrita, inferimos que este conjunto sólo abarca la mitad de esta circunferencia, la que une a los puntos* $(0, 0, 1) = f(t_0, 0)$ *y* $(0, 0, -1) = f(t_0, \pi)$ *y que se proyecta sobre dicha semirecta (ver figura 2.7).*

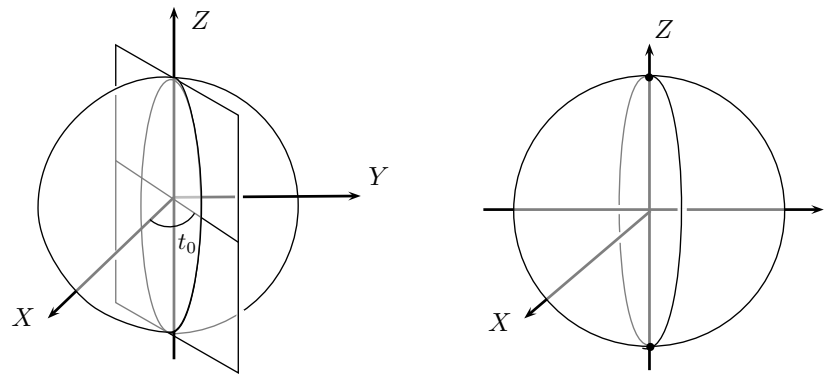


Figura 2.7: El conjunto $f(\{(t_0, s) \mid s \in [0, \pi]\})$ está contenido en la intersección de la esfera $x^2 + y^2 + z^2 = 1$ con el plano $-\mathrm{sen}(t_0)x + \cos(t_0)y = 0$, y sólo consta de una semicircunferencia.

Las propiedades más importantes relacionadas con los conceptos de imagen inversa e imagen directa las dejaremos plasmadas en la siguiente proposición. Dada la sencillez de sus pruebas, éstas quedarán a cargo del lector.

Proposición 2.1 *Sean* $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$*,* $A_\alpha, B, C \subset A$*, y* $D_\alpha, D, E \subset \mathbb{R}^m$*, con* $\alpha \in I$*,* I *un conjunto de índices. Se cumple que:*

1. *si* $D \subset E$*, entonces* $f^{-1}(D) \subset f^{-1}(E)$

2. $f^{-1}\left(\bigcup_{\alpha\in I} D_\alpha\right) = \bigcup_{\alpha\in I} f^{-1}(D_\alpha)$

3. $f^{-1}\left(\bigcap_{\alpha\in I} D_\alpha\right) = \bigcap_{\alpha\in I} f^{-1}(D_\alpha)$

4. $f^{-1}(D^c) = (f^{-1}(D))^c \cap A = A \setminus f^{-1}(D)$

5. *si* $B \subset C$, *entonces* $f(B) \subset f(C)$

6. $f\left(\bigcup_{\alpha\in I} A_\alpha\right) = \bigcup_{\alpha\in I} f(A_\alpha)$

7. $f\left(\bigcap_{\alpha\in I} A_\alpha\right) \subset \bigcap_{\alpha\in I} f(A_\alpha)$ *y si* f *es inyectiva, entonces* $f\left(\bigcap_{\alpha\in I} A_\alpha\right) = \bigcap_{\alpha\in I} f(A_\alpha)$

8. $f(A) \setminus f(B) \subset f(A \setminus B)$ *y si* f *es inyectiva, entonces* $f(A \setminus B) = f(A) \setminus f(B)$

9. $B \subset f^{-1}(f(B))$ *y si* f *es inyectiva, entonces* $B = f^{-1}(f(B))$

10. $f(f^{-1}(D)) \subset D$ *y* $f(f^{-1}(D)) = D$ *si y sólo si* $D \subset f(A)$

11. $f^{-1}(D) = f^{-1}(D \cap f(A))$

Concluimos esta sección mencionando otra forma de representar "geométricamente" a las funciones de $\mathbb{R}^2$ en $\mathbb{R}^2$ y de $\mathbb{R}^3$ en $\mathbb{R}^3$. Esta otra forma se desprende básicamente de las dos maneras "geométricas" en que podemos representar a una pareja (o terna) de números reales. En efecto, como ya hemos mencionado en muchas ocasiones, las parejas o ternas las podemos dibujar, o bien como un punto, o bien como una flecha (o vector). Combinando estas dos representaciones, dada una función $f : A \subset \mathbb{R}^2 \to \mathbb{R}^2$ (o $f : A \subset \mathbb{R}^3 \to \mathbb{R}^3$), la representaremos geométricamente de la siguiente forma: a cada elemento $\hat{x} \in A$ lo dibujaremos como un punto y a $f(\hat{x})$ como un vector "colocado" sobre este punto $\hat{x}$.

Otra forma de decir lo anterior es que en el punto $\hat{x} \in A$ "*sembramos*" el vector $f(\hat{x})$. Tal vez por esta forma más "coloquial" de decirlo es que a este tipo de funciones (y su representación geométrica) también se les conoce con el nombre de "*campos vectoriales*".

Es importante mencionar que en la descripción anterior se está suponiendo que, mientras el sistema coordenado con base en el cual se dibuja a cada $\hat{x} \in A$ es fijo, el que se usa para dibujar a $f(\hat{x})$ tendrá como origen al punto $\hat{x}$ y sus ejes (a menos que se indique lo contrario) serán paralelos a los ejes del primero. Es decir, mientras que el sistema coordenado en el que representamos a los elementos del dominio A no cambia, el que usamos para representar a los valores de f cambiará con cada punto.

Los siguientes ejemplos ilustran esta forma de representar geométricamente a este tipo de funciones.

Ejemplo 2.6 *Considere las siguientes funciones.*

1. $f(x, y) = (x, y)$. *En este caso, como el dominio de* f *es todo* $\mathbb{R}^2$*, cada punto del plano tiene asignado un vector, como se muestra en la figura 2.8* (a).

2. $f(x, y) = (-y, x)$. *Nuevamente, como el dominio de* f *también es todo* $\mathbb{R}^2$*, cada punto del plano tiene asignado un vector, como se muestra en la figura 2.8* (b).

3. $f : \mathbb{R}^3 \setminus \{\hat{0}\} \subset \mathbb{R}^3 \to \mathbb{R}^3$ *dada por*

$$f(x, y, z) = \left(\frac{x}{\sqrt{x^2 + y^2 + z^2}}, \frac{y}{\sqrt{x^2 + y^2 + z^2}}, \frac{z}{\sqrt{x^2 + y^2 + z^2}} \right).$$

En este caso, a cada punto de $\mathbb{R}^3$ *distinto del origen, la función le asigna un vector que tiene la particularidad de ser siempre de norma* 1.

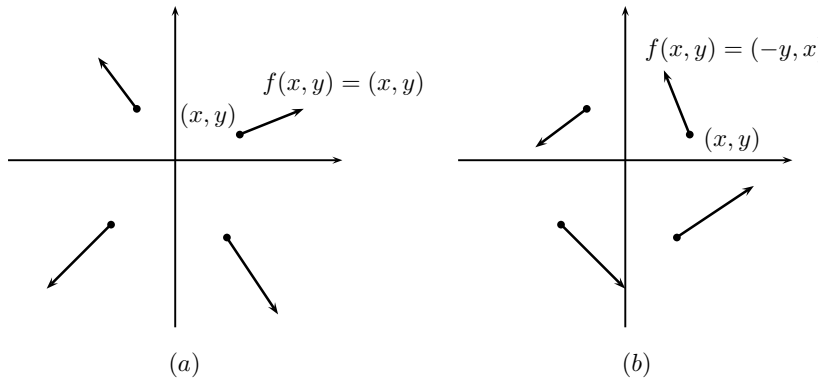


Figura 2.8: Representación geométrica de funciones de $\mathbb{R}^2$ en $\mathbb{R}^2$. La pareja (x, y) se representa por un punto y el valor de f en este punto $(f(x, y))$ por una flecha cuyo punto inicial es el punto (x, y).

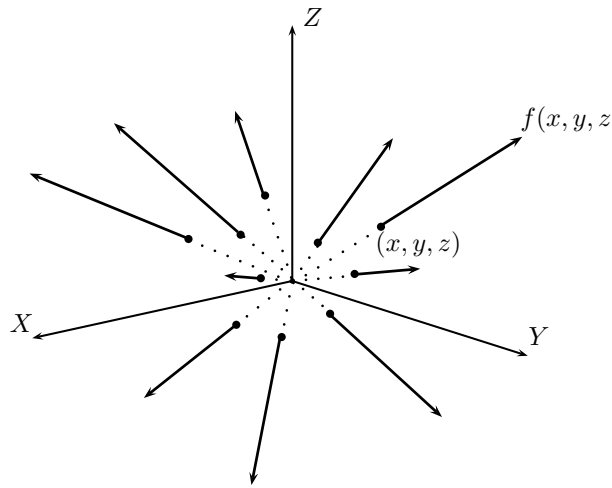


Figura 2.9: Representación geométrica de una función de $\mathbb{R}^3$ en $\mathbb{R}^3$. La terna (x, y, z) se representa por un punto y el valor de f en este punto $(f(x, y, z))$ por una flecha con punto inicial en el punto (x, y, z).

2.2. Límite y continuidad de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$

Como el lector seguramente recordará de su primer curso de cálculo, para el caso de las funciones de $\mathbb{R}$ en $\mathbb{R}$, los conceptos de límite y continuidad están íntimamente relacionados con la idea de cercanía, la cual, tratándose de los números reales, se formaliza a través del concepto de valor absoluto.

Para las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$, los conceptos de límite y continuidad no cambian esencialmente y siguen siendo una expresión de las ideas de cercanía y aproximación.

De esta manera, dado que en $\mathbb{R}^n$ contamos con varias formas de medir la distancia entre sus elementos (o de generalizar el concepto de valor absoluto), en principio tenemos muchas maneras de definir estos conceptos (aunque, como veremos más adelante, y a la luz de las desigualdades de la proposición 1.4 del capítulo 1, todas ellas serán equivalentes).

Así como en el caso de los números reales realizamos nuestro primer acercamiento a la idea de límite (o aproximación) a través del concepto de sucesión, para el caso de $\mathbb{R}^n$ haremos lo mismo.

2.2.1. Sucesiones en $\mathbb{R}^n$

El concepto de sucesión en $\mathbb{R}^n$ (o en un conjunto arbitrario) es análogo al de sucesión en $\mathbb{R}$: una sucesión es una función que a cada natural le asocia un elemento de $\mathbb{R}^n$. Este concepto lo formalizamos en la siguiente

Definición 2.6 *Una sucesión en $\mathbb{R}^n$ es una función $s : \mathbb{N} \to \mathbb{R}^n$. Denotamos por $\hat{x}_k$ a s evaluada en k, es decir, $\hat{x}_k = s(k)$. Para referirnos a la función s, escribiremos $\{\hat{x}_k\}$.*

Como sucede con cualquier función, una sucesión en $\mathbb{R}^n$ queda totalmente determinada si se conoce la regla de asociación, es decir, si conocemos $\hat{x}_k$ para cada $k \in \mathbb{N}$. Dado que $\hat{x}_k \in \mathbb{R}^n$, ésta debe tener n coordenadas (a las cuales denotaremos[2]) por $x_k^{(i)}$, es decir que

[2]En este caso usaremos un superíndice (encerrado entre paréntesis) para denotar la coordenada de un elemento de $\mathbb{R}^n$, dado que en este caso el subíndice denota el término

$$\hat{x}_k = \left(x_k^{(1)}, \ldots, x_k^{(n)}\right).$$

De esta forma, toda sucesión determina (o está determinada por) n sucesiones de números reales, a las cuales conoceremos con el nombre de *sucesiones coordenadas*. Las sucesiones coordenadas tendrán un papel muy importante en los conceptos y resultados que desarrollaremos en esta sección, pues la mayoría de ellos se pueden reducir a los conceptos y resultados análogos del caso real.

Sin duda el concepto más importante con relación a las sucesiones en $\mathbb{R}^n$ es el de convergencia, el cual deseamos que refleje la misma idea de aproximación que se tiene para el caso real.

En términos intuitivos (o geométricos), diremos que una sucesión $\{\hat{x}_k\}$ en $\mathbb{R}^n$ "tiende" (o converge, que es el témino que usaremos) a un punto $\hat{x}_0 \in \mathbb{R}^n$ si los términos de la sucesión están cada vez más cerca de $\hat{x}_0$ conforme el índice k se va haciendo cada vez más grande (o conforme k tiende a infinito, que es la forma en que expresamos esta idea de que "k se va haciendo cada vez más grande").

Como se recordará, en el caso de los números reales esta idea de aproximación o cercanía se formaliza a través del concepto de valor absoluto. Dado que en $\mathbb{R}^n$ contamos con conceptos equivalentes, la definición de convergencia de una sucesión en $\mathbb{R}^n$ será una copia de la definición para sucesiones de números reales.

Definición 2.7 *Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$. Decimos que $\{\hat{x}_k\}$ es convergente (o que converge) si existe $\hat{x}_0 \in \mathbb{R}^n$ tal que para toda cantidad positiva ε existe un índice $N \in \mathbb{N}$ tal que para cualquier otro índice $k \geq N$ se tiene que*

$$\|\hat{x}_k - \hat{x}_0\| < \varepsilon,$$

es decir, si para todo $\varepsilon > 0$ existe $N \in \mathbb{N}$ tal que si $k \geq N$, entonces $\hat{x}_k \in B_\varepsilon(\hat{x}_0)$.

---

de la sucesión.

En este caso decimos que la sucesión $\{\hat{x}_k\}$ *converge a* $\hat{x}_0$*, lo que denotamos como* $\{\hat{x}_k\} \to \hat{x}_0$ *o*

$$\lim_{k\to\infty} \hat{x}_k = \hat{x}_0,$$

Una interpretación geométrica de la definición anterior sería la siguiente: cuando una sucesión $\{\hat{x}_k\}$ converge a un punto $\hat{x}_0$, se tiene que para *cualquier* bola que tomemos con centro en $\hat{x}_0$, sin importar cuán pequeño pueda ser su radio, ésta contiene a casi todos los términos de la sucesión (salvo quizás un número finito de ellos). La figura 2.10 ilustra este hecho en $\mathbb{R}^2$.

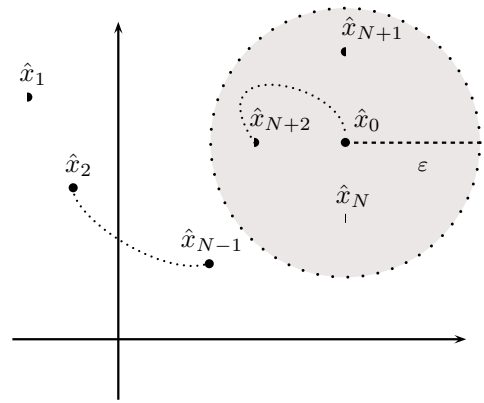


Figura 2.10: Una sucesión $\{\hat{x}_k\}$ converge a un punto $\hat{x}_0$ si para todo $\varepsilon > 0$ existe $N \in \mathbb{N}$ tal que si $k \geq N$, entonces $\hat{x}_k \in B_\varepsilon(\hat{x}_0)$.

Como seguramente el lector habrá pensado, una opción para definir la convergencia de una sucesión en $\mathbb{R}^n$ hubiera sido hacerlo a través de las sucesiones coordenadas, es decir, definir que

$$\left\{\hat{x}_k = \left(x_k^{(1)}, \ldots, x_k^{(n)}\right)\right\}$$

converge si cada sucesión coordenada (de números reales) $\left\{x_k^{(i)}\right\}$ converge (para cada $i \in \{1, \ldots, n\}$). Y el lector habría estado en lo correcto, como lo

mostraremos, dada su importancia, en la primera proposición que formularemos con relación a la convergencia de sucesiones en $\mathbb{R}^n$.

Proposición 2.2 *Sea* $\left\{\hat{x}_k = \left(x_k^{(1)}, \ldots, x_k^{(n)}\right)\right\}$ *una sucesión en* $\mathbb{R}^n$. *La sucesión* $\{\hat{x}_k\}$ *converge si y sólo si la sucesión* $\left\{x_k^{(i)}\right\}$ *converge para cada* $i \in \{1, \ldots, n\}$. *Es decir,* $\{\hat{x}_k\} \to \hat{x}_0 = \left(x_0^{(1)}, \ldots, x_0^{(n)}\right)$ *si y sólo si* $\left\{x_k^{(i)}\right\} \to x_0^{(i)}$ *para cada* $i \in \{1, \ldots, n\}$.

Demostración. Por los incisos 1 y 2 de la proposición 1.4 (del capítulo 1) sabemos que para cualquier $\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ se tiene que

$$|x_i| \leq \|\hat{x}\| \leq |x_1| + \cdots + |x_n|$$

para cada $i \in \{1, \ldots, n\}$, de tal forma que

$$\left|x_k^{(i)} - x_0^{(i)}\right| \leq \|\hat{x}_k - \hat{x}_0\| \leq \left|x_k^{(1)} - x_0^{(1)}\right| + \cdots + \left|x_k^{(n)} - x_0^{(n)}\right| \tag{2.5}$$

para cada $i \in \{1, \ldots, n\}$ y para toda $k \in \mathbb{N}$.

Supongamos ahora que $\{\hat{x}_k\} \to \hat{x}_0 = \left(x_0^{(1)}, \ldots, x_0^{(n)}\right)$. Sabemos entonces que, dado $\varepsilon > 0$, existe un índice $N \in \mathbb{N}$ tal que si $k \geq N$, entonces

$$\|\hat{x}_k - \hat{x}_0\| < \varepsilon,$$

de tal forma que, por la primera desigualdad de 2.5, se tiene (para cada $i \in \{1, \ldots, n\}$) que

$$\left|x_k^{(i)} - x_0^{(i)}\right| \leq \|\hat{x}_k - \hat{x}_0\| < \varepsilon$$

para toda $k \geq N$, de donde concluimos que $\left\{x_k^{(i)}\right\} \to x_0^{(i)}$ (para cada $i \in \{1, \ldots, n\}$).

Si ahora suponemos que $\left\{x_k^{(i)}\right\} \to x_0^{(i)}$ para cada $i \in \{1, \ldots, n\}$, sabemos que dado $\varepsilon > 0$, para $\varepsilon/n > 0$ y para cada $i \in \{1, \ldots, n\}$ existe un índice $N_i \in \mathbb{N}$ tal que

$$\left|x_k^{(i)} - x_0^{(i)}\right| < \frac{\varepsilon}{n}$$

para toda $k \geq N_i$ (y para cada $i \in \{1, \ldots, n\}$). Por tanto, si $N = \text{máx}\{N_1, \ldots, N_n\}$, por la segunda desigualdad de 2.5, si $k \geq N$, como $N \geq N_i$ para cada $i \in \{1, \ldots, n\}$, entonces

$$\begin{aligned} \|\hat{x}_k - \hat{x}_0\| &\leq \left|x_k^{(1)} - x_0^{(1)}\right| + \cdots + \left|x_k^{(n)} - x_0^{(n)}\right| \\ &< \frac{\varepsilon}{n} + \cdots + \frac{\varepsilon}{n} \\ &= \varepsilon, \end{aligned}$$

con lo que concluimos que $\{\hat{x}_k\} \to \hat{x}_0$. ■

Con base en la proposición anterior y dando como un hecho conocido (¡y probado!) que la suma y producto de sucesiones de números reales convergentes también es convergente, obtenemos el siguiente resultado (relacionado con el diferente tipo de operaciones algebraicas que podemos realizar con sucesiones en $\mathbb{R}^n$). Como es de suponerse, la prueba se deja al lector.

Proposición 2.3 *Sean* $\{\hat{x}_k\}, \{\hat{y}_k\}$ *sucesiones en* $\mathbb{R}^n$, *y* $\alpha \in \mathbb{R}$. *Si* $\{\hat{x}_k\} \to \hat{x}_0$ *y* $\{\hat{y}_k\} \to \hat{y}_0$, *entonces:*

1. $\{\hat{x}_k\} + \{\hat{y}_k\} := \{\hat{x}_k + \hat{y}_k\} \to \hat{x}_0 + \hat{y}_0$
2. $\alpha\{\hat{x}_k\} := \{\alpha\hat{x}_k\} \to \alpha\hat{x}_0$
3. $\{\hat{x}_k\} \cdot \{\hat{y}_k\} := \{\hat{x}_k \cdot \hat{y}_k\} \to \hat{x}_0 \cdot \hat{y}_0$
4. *si* $n = 3$, $\{\hat{x}_k\} \times \{\hat{y}_k\} := \{\hat{x}_k \times \hat{y}_k\} \to \hat{x}_0 \times \hat{y}_0$

Hablando de criterios que nos permiten garantizar la convergencia de una sucesión en $\mathbb{R}^n$, seguramente el lector recordará el muy conocido criterio de convergencia de sucesiones de números reales: el afamado *criterio de Cauchy*. Para formular este criterio, antes será necesario introducir el concepto de sucesión de Cauchy en $\mathbb{R}^n$.

Como en el caso de la definición de convergencia en $\mathbb{R}^n$, también podríamos recurrir a las sucesiones coordenadas para "generalizar" el concepto de sucesión de Cauchy a las sucesiones en $\mathbb{R}^n$. Sin embargo, al igual que hicimos

con la convergencia, ya que este concepto también expresa una cierta idea de cercanía, lo definiremos con base al concepto (o a uno de ellos) de distancia que hemos definido en $\mathbb{R}^n$. Después formularemos un resultado que muestra que esta definición es equivalente a que las sucesiones coordenadas también sean de Cauchy.

Definición 2.8 *Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$. Decimos que $\{\hat{x}_k\}$ es de Cauchy si para toda cantidad positiva ε existe un índice $N \in \mathbb{N}$ tal que para cualquier otro par de índices $k, l \geq N$ se tiene que*

$$\|\hat{x}_k - \hat{x}_l\| < \varepsilon,$$

es decir, si para todo $\varepsilon > 0$ existe $N \in \mathbb{N}$ tal que si $k, l \geq N$, entonces $\|\hat{x}_k - \hat{x}_l\| < \varepsilon$.

Como mencionamos antes, la definición anterior resulta equivalente a que las sucesiones coordenadas también sean de Cauchy, lo que plasmamos en la siguiente proposición y cuya prueba, como es de suponerse, queda a cargo del lector.

Proposición 2.4 *Sea $\left\{\hat{x}_k = \left(x_k^{(1)}, \ldots, x_k^{(n)}\right)\right\}$ una sucesión en $\mathbb{R}^n$. La sucesión $\{\hat{x}_k\}$ es de Cauchy si y sólo si la sucesión $\left\{x_k^{(i)}\right\}$ es de Cauchy para cada $i \in \{1, \ldots, n\}$.*

La importancia del concepto de sucesión de Cauchy (en $\mathbb{R}^n$) radica en que dicha propiedad es una condición equivalente a la propiedad de ser convergente (también en $\mathbb{R}^n$), resultado que podemos obtener como un fácil corolario de la proposición anterior y del correspondiente resultado para las sucesiones de números reales.

Corolario 2.1 *Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$. La sucesión $\{\hat{x}_k\}$ es convergente si y sólo si $\{\hat{x}_k\}$ es de Cauchy.*

Es importante mencionar que el resultado anterior se puede probar sin recurrir a la proposición 2.4, prueba que dejamos como un problema para el lector (problema 23).

Otro concepto que resulta muy importante con relación a las sucesiones es el de subsucesión. Con la misma idea bajo la cual se define el concepto de subsucesión de una sucesión de números reales, también se define el concepto de subsucesión de una sucesión en $\mathbb{R}^n$. Es decir, una subsucesión de una sucesión $\{\hat{x}_k\}$ será una nueva sucesión que se construye eligiendo términos de la sucesión original $\{\hat{x}_k\}$, con la única restricción de que los índices de los términos que se elijan vayan "creciendo", es decir, que éstos formen una sucesión de números naturales "creciente" (restricción que no debe extrañarnos, sobre todo si recordamos que lo importante de las sucesiones es lo que sucede con sus términos justo cuando su índice "crece" (o tiende a "infinito")).

La formalización de este concepto la damos en la siguiente

Definición 2.9 *Sean $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$ y $\{k_l\}$ una sucesión de números naturales (es decir, una función que al natural l le asocia el natural k_l). Decimos que $\{\hat{x}_{k_l}\}$ es una subsucesión de $\{\hat{x}_k\}$ si $\{k_l\}$ es una sucesión creciente de números naturales (es decir, que $k_l < k_{l+1}$ para toda $l \in \mathbb{N}$).*

Como dijimos antes, una subsucesión $\{\hat{x}_{k_l}\}$ de una sucesión $\{\hat{x}_k\}$ es a su vez una sucesión, y es importante recalcar que el índice (o variable) de esta nueva sucesión está representado por la letra l; es decir, el décimo término de la sucesión $\{\hat{x}_{k_l}\}$ es el k_{10} término de la sucesión original $\{\hat{x}_k\}$.

También es importante hacer notar que toda sucesión es subsucesión de sí misma; en efecto, si tomamos $k_l = l$ la función identidad de $\mathbb{N}$ en $\mathbb{N}$, que sin duda es creciente, entonces la subsucesión $\{\hat{x}_{k_l}\}$ es $\{\hat{x}_l\}$. Es decir, la sucesión original $\{\hat{x}_k\}$ (en donde la única diferencia es que cambiamos el nombre de su índice (o variable), l por k).

De la idea intuitiva de convergencia es de esperarse que si una sucesión $\{\hat{x}_k\}$ converge al punto $\hat{x}_0 \in \mathbb{R}^n$, entonces cualquier subsucesión de ésta también converja a $\hat{x}_0$. El recíproco de la afirmación anterior también es cierto; es decir, si *todas* las subsucesiones de una sucesión $\{\hat{x}_k\}$ convergen a un mismo punto $\hat{x}_0 \in \mathbb{R}^n$, entonces $\{\hat{x}_k\}$ también converge a $\hat{x}_0$, afirmación que resulta evidente puesto que, como dijimos antes, toda sucesión es subsucesión de sí misma.

Lo interesante es que, aún cuando excluyamos a la sucesión $\{\hat{x}_k\}$ como subsucesión de sí misma, el resultado sigue siendo cierto. Este es un hecho importante y lo dejamos plasmado en la siguiente proposición. Con relación a su prueba, vale la pena mencionar que, como ya va siendo costumbre, hay dos caminos para hacerla: usar la proposición 2.2 y el correspondiente resultado para sucesiones de números reales (que es la que haremos aquí), o hacerla sin usar esta proposición y probarla "directamente" (que es la que dejaremos como un problema para el lector).

Proposición 2.5 *Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$. La sucesión $\{\hat{x}_k\}$ converge al punto $\hat{x}_0 \in \mathbb{R}^n$ si y sólo si cualquier subsucesión $\{\hat{x}_{k_l}\}$ de $\{\hat{x}_k\}$, diferente de $\{\hat{x}_k\}$, también converge a $\hat{x}_0$.*

Demostración. Para entender un poco mejor esta prueba, nótese que la proposición se puede reformular de la siguiente manera: $\{\hat{x}_k\}$ converge al punto $\hat{x}_0 \in \mathbb{R}^n$ si y sólo si para cualquier sucesión creciente de números naturales $\{k_l\}$ tal que $k_l \neq l$ para alguna $l \in \mathbb{N}$, se tiene que $\{\hat{x}_{k_l}\}$ también converge al punto $\hat{x}_0$.

Supongamos que

$$\hat{x}_k = \left(x_k^{(1)}, \ldots, x_k^{(n)}\right) \text{ y } \hat{x}_0 = \left(x_0^{(1)}, \ldots, x_0^{(n)}\right).$$

Si $\{\hat{x}_k\} \to \hat{x}_0$, por la proposición 2.2 sabemos que $\{x_k^{(i)}\} \to x_0^{(i)}$ para cada $i \in \{1, \ldots, n\}$. Ahora, si tomamos cualquier subsucesión $\{\hat{x}_{k_l}\} \neq \{\hat{x}_k\}$, dado que $\{x_{k_l}^{(i)}\}$ es una subsucesión de $\{x_k^{(i)}\}$ (para cada $i \in \{1, \ldots, n\}$), por el correspondiente resultado para sucesiones de números reales sabemos que $\{x_{k_l}^{(i)}\} \to x_0^{(i)}$ (para cada $i \in \{1, \ldots, n\}$) de modo que, nuevamente por la proposición 2.2, tenemos que $\{\hat{x}_{k_l}\} \to \hat{x}_0$.

Por otra parte, si *cualquier* subsucesión $\{\hat{x}_{k_l}\} \neq \{\hat{x}_k\}$ converge a $\hat{x}_0 = \left(x_0^{(1)}, \ldots, x_0^{(n)}\right)$, entonces *cualquier* subsucesión $\left\{x_{k_l}^{(i)}\right\} \neq \left\{x_k^{(i)}\right\}$ converge a $x_0^{(i)}$ (para cada $i \in \{1, \ldots, n\}$). De esta forma, por el correspondiente resultado para sucesiones de números reales, sabemos que $\left\{x_k^{(i)}\right\} \to x_0^{(i)}$

(para cada $i \in \{1, \ldots, n\}$). Por lo tanto, una vez más por la proposición 2.2, tenemos que $\{\hat{x}_k\} \to \hat{x}_0$. ■

El resultado anterior tiene una consecuencia práctica muy importante: si $\{\hat{x}_k\}$ es una sucesión que tiene una subsucesión que no converge, o tiene dos subsucesiones que convergen a puntos diferentes, entonces $\{\hat{x}_k\}$ no es convergente.

Para finalizar con la lista de condiciones necesarias y suficientes (o ambas) para que una sucesión en $\mathbb{R}^n$ sea convergente (sin duda el tema más importante con relación a éstas), retomaremos la interpretación geométrica que dimos al hecho de que una sucesión sea convergente.

Como se recordará, si una sucesión $\{\hat{x}_k\}$ converge a un punto $\hat{x}_0$, entonces cualquier bola con centro en este punto, sin importar su radio, contiene a casi todos los términos de la sucesión, salvo quizás un número finito de ellos. Entre otras cosas, de este hecho se deduce que el conjunto formado por los términos de la sucesión (y que más adelante definiremos formalmente) está ubicado esencialmente alrededor del punto $\hat{x}_0$ y por lo tanto será un conjunto acotado.

El conjunto formado por los términos de una sucesión (al que se le conoce como el rango de la sucesión) es muy importante, y más aún cuando éste está acotado. Por esta razón lo definiremos formalmente y a las sucesiones para las cuales se cumple que su rango es un conjunto acotado, les daremos un nombre particular (que seguramente el lector ya adivina).

Definición 2.10 *Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$. Definimos el rango de la sucesión, que denotamos por $R(\{\hat{x}_k\})$, como el conjunto formado por los términos de la sucesión, es decir*

$$R(\{\hat{x}_k\}) := \{\hat{x}_k \mid k \in \mathbb{N}\}.$$

Definición 2.11 *Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$. Decimos que $\{\hat{x}_k\}$ es una sucesión acotada si $R(\{\hat{x}_k\})$ es un conjunto acotado, es decir, si existe $M > 0$ tal que $\|\hat{x}_k\| \leq M$ para toda $k \in \mathbb{N}$.*

Con base en estas definiciones, la discusión que hicimos previa a ellas se puede resumir en la siguiente

Proposición 2.6 *Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$. Si $\{\hat{x}_k\}$ es una sucesión convergente, entonces $\{\hat{x}_k\}$ es una sucesión acotada.*

Demostración. Sea $\hat{x}_0 \in \mathbb{R}^n$ tal que $\{\hat{x}_k\} \to \hat{x}_0$. De acuerdo con la definición de convergencia, sabemos que para $\varepsilon = 1 > 0$ (o cualquier otra cantidad positiva que se nos ocurra) existe un índice $N \in \mathbb{N}$ tal que si $k \geq N$, entonces $\hat{x}_k \in B_1(\hat{x}_0)$, o equivalentemente, que $\|\hat{x}_k - \hat{x}_0\| < 1$. De esta forma, si tomamos $M = \max\{\|\hat{x}_1\|, \ldots, \|\hat{x}_{N-1}\|, \|\hat{x}_0\| + 1\}$ tenemos que $\|\hat{x}_k\| \leq M$ si $k \in \{1, \ldots, N-1\}$, y si $k \geq N$, entonces $\|\hat{x}_k - \hat{x}_0\| < 1$. Por lo tanto

$$\begin{aligned} \|\hat{x}_k\| &= \|(\hat{x}_k - \hat{x}_0) + \hat{x}_0\| \\ &\leq \|\hat{x}_k - \hat{x}_0\| + \|\hat{x}_0\| \\ &< 1 + \|\hat{x}_0\| \\ &\leq M, \end{aligned}$$

con lo que concluimos que $\|\hat{x}_k\| \leq M$ para toda $k \in \mathbb{N}$, es decir, que $\{\hat{x}_k\}$ es una sucesión acotada. ■

El resultado anterior nos proporciona una consecuencia (o condición) necesaria de las sucesiones convergentes. Suele ser muy útil cuando dicha condición no se cumple pues, nos permite concluir que la sucesión en cuestión no es convergente (como suele suceder con todas las condiciones que son necesarias). Desafortunadamente esta propiedad no es una condición suficiente que garantice la convergencia de una sucesión, como es fácil verificar en el siguiente

Ejemplo 2.7 *Considere la sucesión en $\mathbb{R}^2$ dada por $\{\hat{x}_k = \big((-1)^k, 1/k\big)\}$. Como*

$$\begin{aligned} \|\hat{x}_k\| &= \sqrt{((-1)^k)^2 + (1/k)^2} \\ &= \sqrt{1 + 1/k^2} \\ &\leq \sqrt{2}, \end{aligned}$$

concluimos que $\{\hat{x}_k\}$ es una sucesión acotada. Sin embargo, como $\{x_k^{(1)} = (-1)^k\}$ no es convergente, por la proposición 2.2 tenemos que $\{\hat{x}_k\}$ es una sucesión no convegente.

A pesar del ejemplo anterior, del hecho de que una sucesión $\{\hat{x}_k\}$ sea acotada se puede obtener un resultado muy importante: si bien la sucesión $\{\hat{x}_k\}$ "completa" puede no ser convergente, lo que siempre sucede es que existe al menos una subsucesión de $\{\hat{x}_k\}$ que sí converge. Esta afirmación la podemos verificar con la sucesión del ejemplo anterior, para la cual se tiene que la subsucesión cuyos índices son los números pares, $\{\hat{x}_{2l} = \big((-1)^{2l}, 1/2l\big)\}$, converge al punto $(1,0)$, y que la subsucesión de los índices impares, $\{\hat{x}_{2l-1} = \big((-1)^{2l-1}, 1/(2l-1)\big)\}$, converge al punto $(-1,0)$. Este importante resultado lo dejamos plasmado en el siguiente

Teorema 2.1 *Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$. Si $\{\hat{x}_k\}$ está acotada, entonces existe $\{\hat{x}_{k_l}\}$, subsucesión de $\{\hat{x}_k\}$, tal que $\{\hat{x}_{k_l}\}$ es convergente.*

Demostración. Existen al menos tres pruebas diferentes de este teorema (dos de las cuales se dejan como problema para el lector). La prueba que haremos aquí se basará, como en casos anteriores, en la proposición 2.2 y en el resultado equivalente para sucesiones de números reales.

Procederemos por inducción en n, la dimensión del espacio en el que estamos tomando la sucesión. De esta forma, para $n = 1$ estamos en el caso de sucesiones en $\mathbb{R}$, el cual vamos a dar por probado. Supongamos entonces que el teorema es cierto para sucesiones en $\mathbb{R}^n$ y lo probaremos para una sucesión $\left\{\hat{x}_k = \left(x_k^{(1)}, \ldots, x_k^{(n)}, x_k^{(n+1)}\right)\right\}$ en $\mathbb{R}^{n+1}$ acotada.

Definimos $\left\{\hat{y}_k = \left(x_k^{(1)}, \ldots, x_k^{(n)}\right)\right\}$; por el problema 20 (aplicado en ambos "sentidos") tenemos que $\{\hat{y}_k\}$ es una sucesión acotada en $\mathbb{R}^n$, de tal forma que, por hipótesis de inducción, tiene una subsucesión $\{\hat{y}_{k_l}\}$ que es convergente.

Consideremos ahora la sucesión de números reales $\left\{x_l = x_{k_l}^{(n+1)}\right\}$; dado que $\left\{x_k^{(n+1)}\right\}$ está acotada, entonces $\{x_l\}$ también está acotada, de tal forma

que, por el mismo teorema para sucesiones en $\mathbb{R}$, sabemos que existe una subsucesión $\left\{x_{l_m}^{(n+1)} = x_{k_{l_m}}^{(n+1)}\right\}$ de $\{x_l\}$ que converge.

Ahora, como $\left\{\hat{y}_{k_{l_m}} = \left(x_{k_{l_m}}^{(1)}, \ldots, x_{k_{l_m}}^{(n)}\right)\right\}$ es una subsucesión de $\{\hat{y}_{k_l}\}$, por la proposición 2.5 se tiene que ésta también converge y, por lo tanto (nuevamente por la proposición 2.2, en ambos "sentidos"), obtenemos que la subsucesión

$$\left\{\hat{x}_{k_{l_m}} = \left(x_{k_{l_m}}^{(1)}, \ldots, x_{k_{l_m}}^{(n)}, x_{k_{l_m}}^{(n+1)}\right)\right\}$$

de $\{\hat{x}_k\}$, también converge, afirmación con la que concluye nuestra prueba. ■

2.2.2. Límite

Como sucede con las sucesiones en $\mathbb{R}^n$, los conceptos de límite y continuidad para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ expresan la misma idea de aproximación que expresan los mismos conceptos para el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$, y además se definen de manera completamente análoga a como se hace para este tipo de funciones.

Aún tomando en cuenta lo anterior, es importante hacer notar que ahora que contamos con una clasificación más detallada del tipo de puntos asociados a un conjunto $A \subset \mathbb{R}^n$, si este conjunto es el dominio de una función f, los puntos para los cuales tendrá sentido preguntarse por el límite de f serán justo aquellos que están "pegados" a A, es decir, el tipo de punto al que nos podemos aproximar por medio de puntos diferentes de él y que estén en A (ver inciso (b) del problema 29). Como el lector recordará, estos puntos son aquellos que llamamos puntos de acumulación de A y al conjunto que forman es al que denotamos por A'.

De aquí en adelante, cuando digamos que una sucesión $\{\hat{x}_k\}$ está contenida en en un conjunto A, esto significará que el rango de $\{\hat{x}_k\}$ $(R(\{\hat{x}_k\}))$ está contenido en A. Por otra parte, nótese que, si $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, entonces $\{f(\hat{x}_k)\}$ también es una sucesión (en $\mathbb{R}^m$) y la llamaremos *la sucesión de imágenes (bajo f)*. Una vez dicho lo anterior, damos la siguiente

Definición 2.12 (de límite por sucesiones) *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $\hat{x}_0 \in A'$. Decimos que f tiene límite en $\hat{x}_0$ y que su límite es $\hat{l} \in \mathbb{R}^m$, si para toda sucesión $\{\hat{x}_k\}$ contenida en $A \setminus \{\hat{x}_0\}$ que converge a $\hat{x}_0$ se tiene que la sucesión de imágenes $\{f(\hat{x}_k)\}$ converge a $\hat{l}$. En este caso escribimos que*

$$\lim_{\hat{x} \to \hat{x}_0} f(\hat{x}) = \hat{l}$$

y decimos que $\hat{l}$ es el límite de f en $\hat{x}_0$.

De forma análoga a lo que sucede con las sucesiones en $\mathbb{R}^n$, y como una consecuencia de este hecho, si $f = (f_1, \ldots, f_m)$, la existencia del límite de f en un punto $\hat{x}_0 \in A'$ es una condición necesaria y suficiente para la existencia del límite (en el mismo punto) de cada una de sus funciones coordenadas f_i. Este importante resultado lo dejamos expresado en la siguiente

Proposición 2.7 *Sean $f = (f_1, \ldots, f_m) : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $\hat{x}_0 \in A'$. La función f tiene límite en $\hat{x}_0$ y su límite es $\hat{l} = (l_1, \ldots, l_m) \in \mathbb{R}^m$ si y sólo si la función f_i tiene límite en $\hat{x}_0$ y su límite es l_i, para cada $i \in \{1, \ldots, m\}$.*

La prueba de esta proposición es una consecuencia inmediata de la proposición 2.2 y se deja como un problema para el lector. La proposición anterior tiene la virtud de reducir el problema de determinar la existencia del límite (y en caso de existir, su cálculo) de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ a sólo funciones de $\mathbb{R}^n$ en $\mathbb{R}$. Por esta razón, a continuación daremos una serie de ejemplos en los cuales sólo consideraremos funciones de este último tipo.

Antes de dar los ejemplos, describiremos un procedimiento que consiste en "experimentar" con algunas sucesiones que satisfagan la definición 2.12 y observar qué es lo que sucede con las correspondientes sucesiones de imágenes. Los pasos a seguir después de hacer estos "experimentos" dependerán de sus resultados.

Si para "varias" sucesiones específicas que convergen a un punto $\hat{x}_0$, las correspondientes sucesiones de imágenes siempre convergen a un mismo valor $\hat{l}$, y además el lector "intuye" (intuición que seguramente desarrollará después de calcular muchos límites) que la función debe tener límite,

habrá que hacer una demostración que satisfaga la definición 2.12. Es decir, probar que si $\{\hat{x}_k\}$ es cualquier otra sucesión (totalmente arbitraria) que converge a $\hat{x}_0$, entonces la sucesión de imágenes $\{f(\hat{x}_k)\}$ converge a $\hat{l}$.

Por otra parte, si tenemos la suerte de encontrar sucesiones para las cuales las correspondientes sucesiones de imágenes no convergen, o convergen a valores diferentes, entonces de acuerdo a la definición 2.12 podemos concluir que la función no tiene límite en el punto en cuestión. Como se podrá notar, cuando el problema sea mostrar que una función no tiene límite en un punto, las sucesiones son una herramienta muy útil y muy sencilla de usar.

Ejemplo 2.8 *Determinaremos si las siguientes funciones tienen límite en el punto que se indica.*

1. $f(x,y) = xy/(x^2+y^2)$ *en el punto* $(0,0)$.

   Observe que esta función está definida en $A = \mathbb{R}^2 \setminus \{(0,0)\}$ *y el* $(0,0)$ *es un punto de acumulación de A, de tal forma que sí es válido preguntarse si esta función tiene límite en dicho punto.*

   Tomemos la sucesión $\{\hat{x}_k = (1/k, 0)\}$*; esta sucesión satisface las condiciones de la definición 2.12 ya que* $\hat{x}_k \in A$ *para toda* $k \in \mathbb{N}$*, y* $\{\hat{x}_k\} \to (0,0)$ *en virtud de que* $\{x_k^{(1)} = 1/k\} \to 0$ *y* $\{x_k^{(2)} \equiv 0\} \to 0$ *(la sucesión constante cero). Por otra parte, se tiene que*

$$\begin{aligned} f(\hat{x}_k) &= \frac{(1/k)(0)}{(1/k)^2 + 0^2} \\ &= 0 \end{aligned}$$

   para toda $k \in \mathbb{N}$*, es decir, la sucesión de imágenes* $\{f(\hat{x}_k)\}$ *es la sucesión constante cero, la cual converge a* 0.

   Si ahora consideramos la sucesión $\{\hat{x}_k = (1/k, 1/k)\}$*, también es una sucesión contenida en A para la cual se satisface que* $\{\hat{x}_k\} \to (0,0)$ *y sin embargo*

$$\begin{aligned} f(\hat{x}_k) &= \frac{(1/k)(1/k)}{(1/k)^2 + (1/k)^2} \\ &= \frac{1}{2} \end{aligned}$$

para toda $k \in \mathbb{N}$. *Es decir, ahora la sucesión de imágenes* $\{f(\hat{x}_k)\}$ *es la sucesión constante* $1/2$, *la cual converge a* $1/2$.

Con base en el comportamiento de estas dos sucesiones, podemos concluir que la función no tiene límite en el punto $(0,0)$.

2. $f(x,y) = xy/\sqrt{x^2+y^2}$ *en el punto* $(0,0)$.

Como en el inciso anterior, es fácil ver que si

$$\{\hat{x}_k = (1/k, 0)\}, \ \ \{\hat{x}_k = (0, 1/k)\} \ \ o \ \ \{\hat{x}_k = (1/k, 1/k)\},$$

entonces en todos estos casos se tiene que la correspondientes sucesiones de imágenes satisfacen que $\{f(\hat{x}_k)\} \to 0$.

Más aún, si $\{\hat{x}_k = (1/k, m/k)\}$, *con* $m \in \mathbb{R}$, *en cuyo caso los terminos* $\hat{x}_k$ *satisfacen la ecuación de la recta (en* $\mathbb{R}^2$*)* $y = mx$ *(es decir que la sucesión* $\{\hat{x}_k\}$ *se "aproxima" al* $(0,0)$ *"por" (o "sobre") esta recta), entonces*

$$\begin{aligned} f(\hat{x}_k) &= \frac{(1/k)(m/k)}{\sqrt{(1/k)^2 + (m/k)^2}} \\ &= \frac{m(1/k)^2}{(1/k)\sqrt{1+m^2}} \\ &= \left(\frac{1}{k}\right)\frac{m}{\sqrt{1+m^2}}, \end{aligned}$$

de donde también tenemos que $\{f(\hat{x}_k)\} \to 0$.

Los "experimentos" anteriores nos hacen "sospechar" que la función sí tiene límite (en cuyo caso tendría que ser 0*), y para demostrar que esta afirmación es cierta, recurriremos a una de las desigualdades que el lector probó en el problema 3 del capítulo* 1 *y que resultará muy útil a la hora de hacer demostraciones de límites. En ese problema se prueba que, si* $\hat{x} = (x_1, ..., x_n) \in \mathbb{R}^n$, *entonces* $|x_i| \leq \|\hat{x}\|$ *para cada* $i = 1, ..., n$. *De esta forma, si* $\hat{x} = (x, y)$, *entonces* $|x|, |y| \leq \|\hat{x}\| = \sqrt{x^2+y^2}$ *y por tanto*

$$|xy| \leq \|\hat{x}\|^2.$$

Por esta última desigualdad, si $\hat{x} \neq \hat{0}$*, tenemos que*

$$\begin{aligned} |f(\hat{x})| &= \left| \frac{xy}{\sqrt{x^2+y^2}} \right| \\ &= \frac{|xy|}{\|\hat{x}\|} \\ &\leq \|\hat{x}\|. \end{aligned}$$

Sea ahora $\{\hat{x}_k\}$ *cualquier sucesión contenida en* $A = \mathbb{R}^2 \setminus \{(0,0)\}$ *tal que* $\{\hat{x}_k\} \to (0,0)$*; por la desigualdad anterior, se tiene que*

$$\begin{aligned} |f(\hat{x}_k) - 0| &= |f(\hat{x}_k)| \\ &\leq \|\hat{x}_k\|. \end{aligned}$$

Ahora, como $\{\hat{x}_k\} \to (0,0)$*, por el problema 14 tenemos que* $\{\|\hat{x}_k\|\} \to 0$*, por la desigualdad anterior (aplicando la "ley del sandwich" para sucesiones de números reales) concluimos que la sucesión* $\{|f(\hat{x}_k) - 0|\} \to 0$ *y por el problema 13 tenemos que* $\{f(\hat{x}_k)\} \to 0$*.*

De esta forma, hemos probado que

$$\lim_{(x,y)\to(0,0)} \frac{xy}{\sqrt{x^2+y^2}} = 0.$$

En el ejemplo anterior, inciso (2), usamos la conocida "*ley del sandwich*" para sucesiones de números reales. Este resultado lo podemos "extender" a las funciones de $\mathbb{R}^n$ en $\mathbb{R}$ y va a resultar ser una herramienta muy útil para el cálculo de límites.

Proposición 2.8 *Sean* $f, g, h : A \subset \mathbb{R}^n \to \mathbb{R}$ *y* $\hat{x}_0 \in A'$*. Si existe* $r > 0$ *tal que*

$$f(\hat{x}) \leq h(\hat{x}) \leq g(\hat{x})$$

para toda $\hat{x} \in (B_r(\hat{x}_0) \setminus \{\hat{x}_0\}) \cap A$ *y*

$$\lim_{\hat{x}\to\hat{x}_0} f(\hat{x}) = l = \lim_{\hat{x}\to\hat{x}_0} g(\hat{x}),$$

entonces h tiene límite en $\hat{x}_0$ y

$$\lim_{\hat{x}\to\hat{x}_0} h(\hat{x}) = l.$$

Esta proposición es una consecuencia inmediata de la mencionada "*ley del sandwich*" para sucesiones de números reales y, como es de suponer, su prueba se deja al lector.

Lo que sí vamos a hacer es resaltar el hecho de que la hipótesis relacionada con las desigualdades que deben satisfacer las funciones involucradas sólo se debe de satisfacer en una vecindad del punto en donde se va a tomar el límite, confirmando con ello que este concepto sólo depende del comportamiento "local" de dichas funciones.

Otro hecho muy fácil de probar, pero no por ello menos importante, es la relación del concepto de límite con la aritmética de las funciones. Los resultados que formularemos en la siguiente proposición se deducen (casi todos ellos) de las correspondientes propiedades de sucesiones que quedaron plasmadas en la proposición 2.3. Por esta razón, salvo en el caso de la última afirmación, su prueba se deja al lector.

Proposición 2.9 *Sean $f, g : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $\alpha \in \mathbb{R}$, $\hat{l}, \hat{l}' \in \mathbb{R}^m$ y $\hat{x}_0 \in A'$. Si f y g tienen límite en $\hat{x}_0$ y*

$$\lim_{\hat{x}\to\hat{x}_0} f(\hat{x}) = \hat{l} \qquad y \qquad \lim_{\hat{x}\to\hat{x}_0} g(\hat{x}) = \hat{l}',$$

entonces:

1. *la función $f + g$ tiene límite en $\hat{x}_0$ y además*

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} (f+g)(\hat{x}) &= \lim_{\hat{x}\to\hat{x}_0} f(\hat{x}) + \lim_{\hat{x}\to\hat{x}_0} g(\hat{x}) \\ &= \hat{l} + \hat{l}'\end{aligned}$$

2. *la función αf tiene límite en $\hat{x}_0$ y además*

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} (\alpha f)(\hat{x}) &= \alpha \lim_{\hat{x}\to\hat{x}_0} f(\hat{x}) \\ &= \alpha\hat{l}\end{aligned}$$

3. la función $f \cdot g$ tiene límite en $\hat{x}_0$ y además

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0}(f\cdot g)(\hat{x}) &= \left(\lim_{\hat{x}\to\hat{x}_0} f(\hat{x})\right)\cdot\left(\lim_{\hat{x}\to\hat{x}_0} g(\hat{x})\right)\\ &= \hat{l}\cdot\hat{l}'\end{aligned}$$

4. si $m = 3$, la función $f \times g$ tiene límite en $\hat{x}_0$ y además

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0}(f\times g)(\hat{x}) &= \left(\lim_{\hat{x}\to\hat{x}_0} f(\hat{x})\right)\times\left(\lim_{\hat{x}\to\hat{x}_0} g(\hat{x})\right)\\ &= \hat{l}\times\hat{l}'\end{aligned}$$

5. si $m = 1$, $B = \{\hat{x} \in A \mid g(\hat{x}) \neq 0\}$ y $\hat{l}' \neq 0$, entonces $\hat{x}_0 \in B'$ y además

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0}\left(\frac{f}{g}\right)(\hat{x}) &= \frac{\lim_{\hat{x}\to\hat{x}_0} f(\hat{x})}{\lim_{\hat{x}\to\hat{x}_0} g(\hat{x})}\\ &= \frac{\hat{l}}{\hat{l}'}\end{aligned}$$

Demostración. (inciso 5). Primero probaremos que $\hat{x}_0 \in B'$. Como $\hat{x}_0 \in A'$, de acuerdo con el inciso (b) del problema 29 sabemos que existe una sucesión $\{\hat{x}_k\}$ contenida en $A \setminus \{\hat{x}_0\}$ tal que $\{\hat{x}_k\}$ converge a $\hat{x}_0$. De la definición de límite tenemos entonces que la sucesión (de números reales) $\{g(\hat{x}_k)\}$ converge a $\hat{l}' \neq 0$, de tal forma que, para $\varepsilon = \left|\hat{l}'\right| > 0$, sabemos que existe $N \in \mathbb{N}$ tal que, si $k \geq \mathbb{N}$, entonces

$$\left|g(\hat{x}_k) - \hat{l}'\right| < \varepsilon = \left|\hat{l}'\right|.$$

Nótese que en la desigualdad anterior no puede suceder que $g(\hat{x}_k) = 0$ y por tanto podemos concluir que, si $k \geq \mathbb{N}$, entonces $g(\hat{x}_k) \neq 0$. De aquí tenemos que $\hat{x}_k \in B \setminus \{\hat{x}_0\}$ para toda $k \geq \mathbb{N}$. Por tanto, si hacemos $\hat{y}_l = \hat{x}_{l+N}$, entonces $\{\hat{y}_l\}$ es una sucesión en $B \setminus \{\hat{x}_0\}$ que, además de ser una subsucesión de la sucesión $\{\hat{x}_k\}$ (tomando $k_l = l + N$), se tiene que también

converge a $\hat{x}_0$. De este modo, nuevamente por el inciso (b) del problema 29, concluimos que $\hat{x}_0 \in B'$.

Esto prueba que $\hat{x}_0$ es punto de acumulación del dominio de la función f/g; por otra parte, que esta función tiene límite en $\hat{x}_0$, y que su límite es el cociente de los límites de f y g, es una consecuencia inmediata del correspondiente resultado para el cociente de sucesiones de números reales, que aquí daremos por probado. ■

Como el lector habrá notado, en las operaciones entre funciones mencionadas en la proposición anterior no se incluye a la composición de funciones. El motivo es que el posible resultado que se podría formular para esta operación no resulta cierto: si $g : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $\hat{x}_0 \in A'$, $f : D \subset \mathbb{R}^m \to \mathbb{R}^k$, $\hat{l} \in D'$ y $\hat{l}' \in \mathbb{R}^k$ son tales que:

1. $\lim_{\hat{x}\to\hat{x}_0} g(\hat{x}) = \hat{l}$ y $\lim_{\hat{y}\to\hat{l}} f(\hat{y}) = \hat{l}'$, y
2. $\hat{x}_0 \in (g^{-1}(D))'$ (esta condición es para garantizar que $\hat{x}_0$ sea punto de acumulación del dominio de la función $f \circ g$),

no es posible asegurar que la función $f \circ g$ tenga límite en el punto $\hat{x}_0$, como lo mostraremos en el siguiente

Ejemplo 2.9 *Sean $g(x, y) = x$ para $(x, y) \in \mathbb{R}^2$ y $f(t) = 0$ si $t \neq 0$ y $f(0) = 1$.*

Lo primero que se tiene que observar en este ejemplo es que el dominio de la función g está dado por $A = \mathbb{R}^2$ y que $(0, 0) \in A'$.

Tomemos ahora $\{\hat{x}_k\}$ cualquier sucesión contenida en A tal que $\{\hat{x}_k = (x_k, y_k)\} \to (0, 0)$. Por la proposición 2.2 sabemos que $\{x_k\} \to 0$, de tal forma que, como $g(\hat{x}_k) = x_k$ para toda $k \in \mathbb{N}$, entonces $\{g(\hat{x}_k)\} \to 0$. Por lo tanto concluimos que $\lim_{(x,y)\to(0,0)} g(x, y) = 0$.

Observemos que la función f está definida para toda $t \in \mathbb{R}$ y que

$$\lim_{t\to 0} f(t) = 0.$$

Por tanto, la función $f \circ g$ también está definida para toda $(x, y) \in A$, de modo que sí es válido preguntar si esta función tiene límite en el punto $(0, 0)$.

Afirmamos que dicho límite no existe; en efecto, si tomamos la sucesión $\{\hat{x}_k = (0, 1/k)\}$*, se tiene que* $\{\hat{x}_k\} \to (0,0)$*. Por otra parte,* $g(\hat{x}_k) = g(0,1/k) = 0$*, de modo que* $(f \circ g)(\hat{x}_k) = 1$ *para toda* $k \in \mathbb{N}$ *y por tanto* $\{(f \circ g)(\hat{x}_k)\} \to 1$*. Si ahora elegimos la sucesión* $\{\hat{y}_k = (1/k, 0)\}$*, entonces* $\{\hat{y}_k\} \to (0,0)$ *y* $g(\hat{y}_k) = g(1/k, 0) = 1/k \neq 0$*, de modo que* $(f \circ g)(\hat{y}_k) = 0$ *para toda* $k \in \mathbb{N}$ *y por tanto* $\{(f \circ g)(\hat{y}_k)\} \to 0$*. De esta forma, concluimos que la función* $f \circ g$ *no tiene límite en el punto* $(0,0)$*.*

Más adelante, una vez que hayamos introducido el concepto de continuidad, retomaremos el problema de formular un resultado que relacione la composición de funciones y el concepto de límite.

Si bien es cierto que definir el concepto de límite a través de sucesiones resulta ser muy intuitivo (y muy útil, cuando se trata de probar que una función *no* tiene límite en un punto), también es cierto que esta definición puede resultar un poco "*complicada*" de usar cuando se trata de probar que una función *sí* tiene límite en un punto.

Lo que ahora vamos a hacer será introducir otra forma de definir el concepto de límite (la muy conocida definición $\varepsilon - \delta$) que, entre otras muchas virtudes, resultará ser más sencilla de usar cuando se quiera probar que una función *sí* tiene límite en un punto. Como es de suponerse, probaremos que ambas definiciones resultan ser equivalentes.

Definición 2.13 $(\varepsilon - \delta)$ *Sean* $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ *y* $\hat{x}_0 \in A'$*. Decimos que* f *tiene límite en* $\hat{x}_0$ *y que su límite es* $\hat{l} \in \mathbb{R}^m$*, si para toda* $\varepsilon > 0$ *existe* $\delta > 0$ *tal que si* $\|\hat{x} - \hat{x}_0\| < \delta$ *(y* $\hat{x} \in A \setminus \{\hat{x}_0\}$*), entonces* $\left\|f(\hat{x}) - \hat{l}\right\| < \varepsilon$*. Es decir, si para toda bola de radio* $\varepsilon > 0$ *(con centro en* $\hat{l}$*) existe una bola de radio* $\delta > 0$ *(con centro en* $\hat{x}_0$*) tal que, si* $\hat{x} \in B_\delta(\hat{x}_0) \cap (A \setminus \{\hat{x}_0\})$*, entonces* $f(\hat{x}) \in B_\varepsilon(\hat{l})$ *(o* $f(B_\delta(\hat{x}_0) \cap (A \setminus \{\hat{x}_0\})) \subset B_\varepsilon(\hat{l})$*).*

Una de las ventajas de la definición anterior es que se puede ilustrar geométricamente, como se muestra en la figura 2.11.

Probar que la definición anterior es equivalente a la definición 2.12 es sin duda algo que tenemos que hacer, pero antes mostraremos por medio de un ejemplo la conveniencia de usar esta definición, sobre todo cuando pretendamos demostrar que una función sí tiene límite.

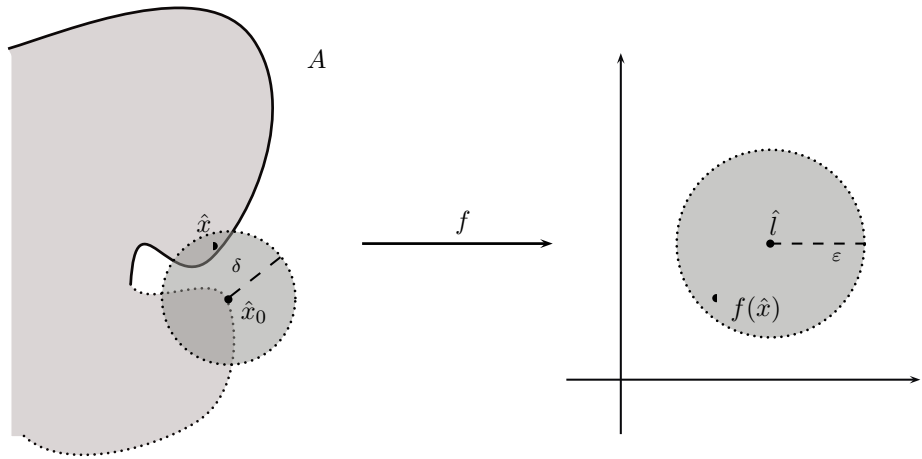


Figura 2.11: La función f tiene límite en $\hat{x}_0 \in A'$ y su límite es $\hat{l} \in \mathbb{R}^m$ si para toda bola de radio $\varepsilon > 0$ (con centro en $\hat{l}$) existe una bola de radio $\delta > 0$ (con centro en $\hat{x}_0$) tal que, si $\hat{x} \in B_\delta(\hat{x}_0) \cap (A \setminus \{\hat{x}_0\})$, entonces $f(\hat{x}) \in B_\varepsilon(\hat{l})$.

Ejemplo 2.10 *Considere la función*

$$f(x,y,z) = \frac{xy^2}{\sqrt{x^4+y^4+z^2}+z^4},$$

que está definida para todo $(x,y,z) \in A = \mathbb{R}^3 \setminus \{(0,0,0)\}$. *Mostraremos, usando la definición 2.13, que esta función tiene límite en el* $\hat{0} = (0,0,0) \in A'$ *y que dicho límite es* 0.

Sea pues una cantidad $\varepsilon > 0$*; nuestra tarea es mostrar que existe una cantidad* $\delta > 0$ *(que en general dependerá de la cantidad* ε *dada) para la cual se satisfaga que, si* $\|\hat{x} - \hat{0}\| < \delta$ *(y* $\hat{x} \in A \setminus \{\hat{0}\}$*), entonces* $|f(\hat{x}) - 0| < \varepsilon$.

Para lograr esto, como seguramente el lector recordará de sus cursos anteriores de cálculo, lo que hay que buscar es la forma de acotar la cantidad $|f(\hat{x}) - 0| = |f(\hat{x})|$ *en términos de la cantidad* $\|\hat{x} - \hat{0}\| = \|\hat{x}\|$*, y para ello habrá que echar mano de todo el acervo de desigualdades de las que dispo-*

nemos. Por ejemplo, recordemos que todo número elevado a una potencia par siempre es no negativo, de tal forma que

$$\sqrt{x^4+y^4+z^2} \leq \sqrt{x^4+y^4+z^2}+z^4$$

y por lo tanto

$$\frac{1}{\sqrt{x^4+y^4+z^2}+z^4} \leq \frac{1}{\sqrt{x^4+y^4+z^2}},$$

lo cual es cierto para toda $(x,y,z) \in A$.

Si por otra parte observamos que la cantidad $\sqrt{x^4+y^4+z^2}$ *es la norma (euclidiana) del vector* (x^2,y^2,z) *y que esta última siempre es mayor o igual que el valor absoluto de cualquiera de sus coordenadas (problema 3 del capítulo* 1*), es decir, que*

$$y^2 = \left|y^2\right| \leq \left\|(x^2,y^2,z)\right\| = \sqrt{x^4+y^4+z^2},$$

entonces concluimos que

$$\frac{y^2}{\sqrt{x^4+y^4+z^2}} \leq 1$$

también para toda $(x,y,z) \in A$.

Por tanto, reuniendo las desigualdades anteriores, tenemos que

$$\begin{aligned}
|f(\hat{x})-0| &= |f(\hat{x})| \\
&= \left|\frac{xy^2}{\sqrt{x^4+y^4+z^2}+z^4}\right| \\
&\leq \frac{|x|\,y^2}{\sqrt{x^4+y^4+z^2}} \\
&= |x|\,\frac{y^2}{\sqrt{x^4+y^4+z^2}} \\
&\leq |x| \\
&\leq \|(x,y,z)\| \\
&= \left\|\hat{x}-\hat{0}\right\|.
\end{aligned}$$

Si tomamos $\delta = \varepsilon$ *y* $\hat{x} \in A$ *es tal que* $\|\hat{x} - \hat{0}\| < \delta = \varepsilon$, *entonces, como* $|f(\hat{x}) - 0| \leq \|\hat{x} - \hat{0}\|$, *tenemos que* $|f(\hat{x}) - 0| < \varepsilon$.

De esta forma mostramos que se verifica la definición 2.13 y concluimos nuestro ejemplo.

Una vez hecho el ejemplo anterior, mostraremos que las dos definiciones de límite que hemos dado son equivalentes, lo cual dejaremos formulado en el siguiente

Teorema 2.2 *Sean* $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $\hat{x}_0 \in A'$ *y* $\hat{l} \in \mathbb{R}^m$. *Si* f *y* $\hat{l}$ *satisfacen la definición 2.13, entonces satisfacen la definición 2.12. Recíprocamente, si* f *y* $\hat{l}$ *satisfacen la definición 2.12, entonces satisfacen la definición 2.13.*

Demostración. Supongamos que f y $\hat{l}$ satisfacen la definición 2.13. Para probar que satisfacen la definición 2.12, tomemos $\{\hat{x}_k\}$ una sucesión contenida en $A \setminus \{\hat{x}_0\}$ tal que $\{\hat{x}_k\} \to \hat{x}_0$; mostraremos que $\{f(\hat{x}_k)\} \to \hat{l}$.

Para probar esto último, tomamos cualquier $\varepsilon > 0$; de la definición 2.13 sabemos que existe $\delta > 0$ tal que si $\hat{x} \in B_\delta(\hat{x}_0) \cap (A \setminus \{\hat{x}_0\})$, entonces $f(\hat{x}) \in B_\varepsilon(\hat{l})$.

Ahora, dado que $\{\hat{x}_k\} \to \hat{x}_0$, para esta $\delta > 0$ sabemos que existe $N \in \mathbb{N}$ tal que, si $k \geq N$, entonces $\|\hat{x}_k - \hat{x}_0\| < \delta$, es decir $\hat{x}_k \in B_\delta(\hat{x}_0) \cap (A \setminus \{\hat{x}_0\})$. Así, por la propiedad que tiene δ, tenemos que $f(\hat{x}_k) \in B_\varepsilon(\hat{l})$, es decir que $\left\|f(\hat{x}_k) - \hat{l}\right\| < \varepsilon$ para toda $k \geq N$, con lo que concluimos que $\{f(\hat{x}_k)\} \to \hat{l}$.

Supongamos ahora que f y $\hat{l}$ satisfacen la definición 2.12. En esta parte de la prueba procederemos por el método de la contrapuesta, es decir, que f y $\hat{l}$ no satisfacen la definición 2.13. Esto significa que existe una cantidad $\varepsilon > 0$ tal que, para cualquier cantidad $\delta > 0$ que se tome, siempre existe $\hat{x} \in B_\delta(\hat{x}_0) \cap (A \setminus \{\hat{x}_0\})$ con la propiedad de que $f(\hat{x}) \notin B_\varepsilon(\hat{l})$, es decir que $\left\|f(\hat{x}) - \hat{l}\right\| \geq \varepsilon$.

Aplicando lo anterior para cada $k \in \mathbb{N}$, podemos tomar $\hat{x}_k \in A \setminus \{\hat{x}_0\}$ tal que $\|\hat{x}_k - \hat{x}_0\| < 1/k$ y $\left\|f(\hat{x}_k) - \hat{l}\right\| \geq \varepsilon$.

De esta forma obtenemos una sucesión $\{\hat{x}_k\}$ contenida en $A \setminus \{\hat{x}_0\}$ tal que $\{\hat{x}_k\} \to \hat{x}_0$ (problema 13) para la cual se tiene que $\{f(\hat{x}_k)\}$ no converge

a $\hat{l}$ (puesto que $\left\| f(\hat{x}_k) - \hat{l} \right\| \geq \varepsilon > 0$ para toda $k \in \mathbb{N}$). Como esta propiedad es justo la negación de nuestra hipótesis, con esto concluimos la prueba de la segunda parte del teorema. ■

2.2.3. Continuidad

Como en el caso de las funciones de $\mathbb{R}$ en $\mathbb{R}$, una de las primeras aplicaciones del concepto de límite está en la formalización del concepto de continuidad de una función.

Dada una función $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, la idea intuitiva de que ésta sea continua consiste en que si $\hat{x}, \hat{y}$ son elementos de su dominio que están "cercanos", entonces sus valores bajo f ($f(\hat{x})$ y $f(\hat{y})$) están "cercanos". Esta idea intuitiva se puede simplificar si dejamos fijo uno de estos puntos, que ahora llamaremos $\hat{x}_0$, y decimos que la función f es "continua" en $\hat{x}_0 \in A$ si para todo $\hat{x} \in A$ que está "cerca" de $\hat{x}_0$ se tiene que $f(\hat{x})$ está "cerca" de $f(\hat{x}_0)$.

Como seguramente el lector ya está intuyendo, esta última expresión se puede formalizar usando el concepto de límite. En efecto, podemos decir que una función f es "continua" en $\hat{x}_0 \in A$ si tiene límite en este punto y su límite es $f(\hat{x}_0)$ (el valor de f en $\hat{x}_0$). Es decir, si

$$\lim_{\hat{x} \to \hat{x}_0} f(\hat{x}) = f(\hat{x}_0). \tag{2.6}$$

Aún cuando todo parezca indicar que hemos logrado una "buena" definición de continuidad, es importante precisar algunos aspectos. El primero de ellos es que el concepto de continuidad se definirá para un punto y éste siempre tendrá que ser un elemento de su dominio, es decir, un punto para el cual la función f está definida. El segundo aspecto tiene que ver con el uso del concepto de límite; como se recordará, el límite de una función sólo se define para los puntos de acumulación del dominio de ésta, lo que puede ser un obstáculo ya que, dado $A \subset \mathbb{R}^n$, en general no es cierto que todo elemento de A tiene que ser un punto de acumulación de A (es decir, $A \not\subseteq A'$).

Si dado $\hat{x}_0 \in A$ tenemos la suerte de que éste sea un punto de acumulación de A ($\hat{x}_0 \in A'$), usar la identidad 2.6 para decir que f es continua en $\hat{x}_0$ es sin duda la forma más adecuada de hacerlo. Sólo restaría analizar el caso en que $\hat{x}_0$ no fuera un punto de acumulación de A (aún cuando sí pertenezca a A). Nótese que en este caso, $\hat{x}_0$ sería un punto aislado de A, es decir, un punto de A para el cual existe $r > 0$ con la propiedad de que $B_r(\hat{x}_0) \cap A = \{\hat{x}_0\}$ (ver definición 1.10), lo que significa que el único punto de A que realmente está "cerca" de $\hat{x}_0$ es ¡él mismo!

Pareciera un poco "inútil" tratar de incluir a los puntos aislados del dominio de una función en el tipo de puntos para los cuales se pueda dar la definición de continuidad (lo que por otra parte, tampoco tendría nada de "insensato"). Sin embargo, aún sin buscar lo anterior, si escribimos la identidad 2.6 usando la definición 2.13, notaremos que estos puntos se pueden incluir de manera "natural". En efecto, de acuerdo con esta definición, que se satisfaga la identidad 2.6 va a significar que, para cualquier cantidad $\varepsilon > 0$ existe una cantidad $\delta > 0$ tal que, si $\hat{x} \in A \setminus \{\hat{x}_0\}$ tiene la propiedad de que $\|\hat{x} - \hat{x}_0\| < \delta$, entonces $\|f(\hat{x}) - f(\hat{x}_0)\| < \varepsilon$, es decir que, si $\hat{x} \in B_\delta(\hat{x}_0) \cap (A \setminus \{\hat{x}_0\})$, entonces $f(\hat{x}) \in B_\varepsilon(f(\hat{x}_0))$.

Si ahora observamos que, como ahora sí estamos seguros de que $\hat{x}_0 \in A$, y que las condiciones anteriores se siguen cumpliendo aún cuando $\hat{x}$ sea igual a $\hat{x}_0$, todo parece indicar que, independientemente de qué tipo de punto sea $\hat{x}_0$ con respecto de A (de acumulación o aislado), la siguiente definición es la mejor opción para expresar el hecho de que una función sea continua en un punto de su dominio.

Definición 2.14 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $\hat{x}_0 \in A$. Decimos que f es continua en $\hat{x}_0$ si para cualquier cantidad $\varepsilon > 0$ existe una cantidad $\delta > 0$ tal que, si $\hat{x} \in A$ tiene la propiedad de que $\|\hat{x} - \hat{x}_0\| < \delta$, entonces $\|f(\hat{x}) - f(\hat{x}_0)\| < \varepsilon$. Es decir que, si $\hat{x} \in B_\delta(\hat{x}_0) \cap A$, entonces $f(\hat{x}) \in B_\varepsilon(f(\hat{x}_0))$ (o equivalentemente, que $f(B_\delta(\hat{x}_0) \cap A) \subset B_\varepsilon(f(\hat{x}_0))$).*

Todas las observaciones y afirmaciones que hicimos para "deducir" la definición anterior las dejaremos plasmadas en la siguiente proposición, y aprovechando el teorema 2.2, incluiremos una forma equivalente de expresar

la continuidad de una función en un punto, en términos de sucesiones (que bien podría llamarse la "*definición de continuidad por sucesiones*").

Proposición 2.10 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $\hat{x}_0 \in A$.*

1. *Si $\hat{x}_0$ es un punto aislado de A, entonces f es continua en $\hat{x}_0$.*
2. *Si $\hat{x}_0$ es un punto de acumulación de A, se satisface que f es continua en $\hat{x}_0$ si y sólo si f tiene límite en $\hat{x}_0$ y además*

$$\lim_{\hat{x}\to\hat{x}_0} f(\hat{x}) = f(\hat{x}_0).$$

3. *La función f es continua en $\hat{x}_0$ si y sólo si para toda sucesión $\{\hat{x}_k\}$ contenida en A que converge a $\hat{x}_0$ se tiene que la sucesión $\{f(\hat{x}_k)\}$ converge a $f(\hat{x}_0)$.*

La demostración de esta proposición es muy sencilla y se deja como un problema para el lector.

Otro hecho (que sin duda también resultará evidente para el lector) es la relación que existe entre la continuidad de una función $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ en un punto $\hat{x}_0 \in A$ y la de sus funciones coordenadas (en el mismo punto).

En efecto, si $f = (f_1, \ldots, f_m)$, es de esperarse que f será continua en $\hat{x}_0 \in A$ si y sólo si f_i es continua en $\hat{x}_0$ para cada $i \in \{1, \ldots, m\}$, propiedad que, a pesar de su "predecibilidad", es muy importante y por lo mismo la dejamos expresada en la siguiente

Proposición 2.11 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $\hat{x}_0 \in A$. La función f es continua en $\hat{x}_0$ si y sólo si f_i es continua en $\hat{x}_0$, para cada $i \in \{1, \ldots, m\}$.*

Las propiedades del concepto de continuidad relacionadas con la aritmética de las funciones, las dejaremos expresadas en la siguiente proposición. Dada la cercanía de este concepto con el de límite, casi todas las afirmaciones de esta proposición (salvo por el inciso 6) serán una consecuencia inmediata de las correspondientes afirmaciones de la proposición 2.9, en virtud de lo cual su prueba, una vez más, quedará a cargo del lector (incluyendo el inciso 6).

Proposición 2.12 *Sean $f, g : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $\alpha \in \mathbb{R}$ y $\hat{x}_0 \in A$. Si f y g son continuas en $\hat{x}_0$, entonces:*

1. *la función $f + g$ es continua en $\hat{x}_0$*
2. *la función αf es continua en $\hat{x}_0$*
3. *la función $f \cdot g$ es continua en $\hat{x}_0$*
4. *si $m = 3$, la función $f \times g$ es continua en $\hat{x}_0$*
5. *si $m = 1$ y $g(\hat{x}_0) \neq 0$, la función f/g es continua en $\hat{x}_0$*
6. *si $\hat{y}_0 = f(\hat{x}_0) \in D \subset \mathbb{R}^m$ y $h : D \subset \mathbb{R}^m \to \mathbb{R}^k$ es continua en $\hat{y}_0$, entonces $h \circ f$ es continua en $\hat{y}_0$.*

Como el lector habrá notado, dentro de las operaciones entre funciones que acabamos de enlistar en la proposición anterior, ahora sí incluimos a la operación composición (a diferencia de la proposición 2.9). La razón de esto, es que el concepto de continuidad sí se "comporta" bien con esta operación (a diferencia del concepto de límite, como vimos en el ejemplo 2.9). Lo que ahora queremos destacar es que hay un resultado un poco más general para el cual no se requiere que ambas funciones (las que se vayan a componer) tengan que ser continuas en los correspondientes puntos (como se requiere en el inciso 6 de la proposición anterior). Este resultado es muy importante (y muy útil) y lo dejaremos formulado en la siguiente

Proposición 2.13 *Sean $g : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $f : D \subset \mathbb{R}^m \to \mathbb{R}^k$, $\hat{x}_0 \in \mathbb{R}^n$ y $\hat{l} \in \mathbb{R}^m$ tales que $\hat{l} \in D$ y $\hat{x}_0 \in (g^{-1}(D))' \subset A'$. Si $\lim_{\hat{x}\to\hat{x}_0} g(\hat{x}) = \hat{l}$ y f es continua en $\hat{l}$, entonces $f \circ g$ tiene límite en $\hat{x}_0$ y además*

$$\lim_{\hat{x}\to\hat{x}_0} (f \circ g)(\hat{x}) = f\left(\hat{l}\right) = f\left(\lim_{\hat{x}\to\hat{x}_0} g(\hat{x})\right).$$

Demostración. Usaremos la definición $\varepsilon - \delta$ de límite para hacer esta demostración. Sea $\varepsilon > 0$ arbitraria. De la definición de continuidad sabemos que existe $\delta' > 0$ tal que

$$\text{si } \hat{y} \in B_{\delta'}\left(\hat{l}\right) \cap D, \text{ entonces } f(\hat{y}) \in B_{\varepsilon}\left(f\left(\hat{l}\right)\right) \tag{2.7}$$

y de la definición de límite tenemos que existe $\delta > 0$ tal que

$$\text{si } \hat{x} \in B_\delta(\hat{x}_0) \cap (A \setminus \{\hat{x}_0\}), \text{ entonces } g(\hat{x}) \in B_{\delta'}\left(\hat{l}\right). \tag{2.8}$$

Afirmamos que, si $\hat{x} \in B_\delta(\hat{x}_0) \cap (B \setminus \{\hat{x}_0\})$, entonces $(f \circ g)(\hat{x}) \in B_\varepsilon\left(f\left(\hat{l}\right)\right)$ (ver figura 2.12). En efecto, dado que $g^{-1}(D) \subset A$, entonces

$$B_\delta(\hat{x}_0) \cap (g^{-1}(D) \setminus \{\hat{x}_0\}) \subset B_\delta(\hat{x}_0) \cap (A \setminus \{\hat{x}_0\}).$$

de tal forma que, si $\hat{x} \in B_\delta(\hat{x}_0) \cap (g^{-1}(D) \setminus \{\hat{x}_0\})$, entonces $\hat{x} \in B_\delta(\hat{x}_0) \cap (A \setminus \{\hat{x}_0\})$. Por lo tanto, de 2.8 se siguee que $g(\hat{x}) \in B_{\delta'}\left(\hat{l}\right)$ y por 2.7 concluimos que $(f \circ g)(\hat{x}) = f(g(\hat{x})) \in B_\varepsilon\left(f\left(\hat{l}\right)\right)$. ■

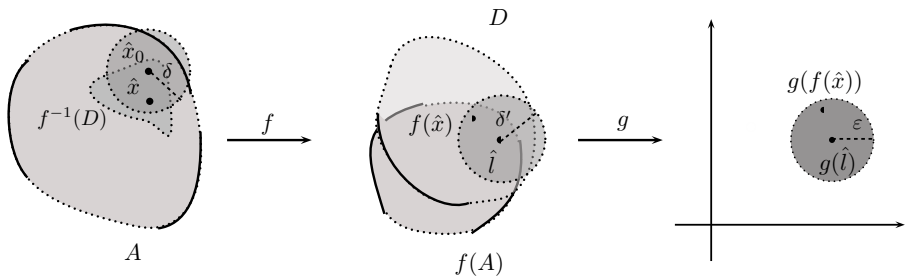


Figura 2.12: "Prueba geométrica" de la proposición 2.13.

Con frecuencia se suele decir "coloquialmente" que el resultado anterior nos asegura que "el límite se puede meter dentro (o a través) de una función continua". Esta "propiedad" es de mucha utilidad cuando nos enfrentamos al problema de mostrar que una función tiene límite, como lo haremos ver en el siguiente

Ejemplo 2.11 *Sea*

$$h(x, y) = \cos\left(\frac{x^2 y}{x^2 + y^2}\right).$$

Mostraremos que esta función tiene límite en el punto $(0,0)$ *y calcularemos cuál es su valor.*

Dado que la función $f(t) = \cos(t)$ *es continua en* $\mathbb{R}$*, por la proposición 2.13 bastará mostrar que la función*

$$g(x,y) = \frac{x^2 y}{x^2 + y^2}$$

tiene límite en el punto $(0,0)$ *y calcular cuánto vale éste. Para ello, obsérvese que, como*

$$\begin{aligned} \left|x^2 y\right| &= x^2 \left|y\right| \\ &\leq \left(x^2 + y^2\right) \|(x,y)\|, \end{aligned}$$

entonces se tiene que

$$\left|\frac{x^2 y}{x^2 + y^2}\right| \leq \|(x,y)\|$$

para toda $(x,y) \neq (0,0)$ *y por lo tanto, que*

$$\begin{aligned} |g(x,y) - 0| &= \left|\frac{x^2 y}{x^2 + y^2}\right| \\ &\leq \|(x,y)\| \\ &= \|(x,y) - (0,0)\|. \end{aligned}$$

De esta forma, dado $\varepsilon > 0$*, si tomamos* $\delta = \varepsilon$ *y* $(x,y) \in \mathbb{R}^2$ *tal que* $0 < \|(x,y) - (0,0)\| < \delta$*, entonces*

$$|g(x,y) - 0| \leq \|(x,y) - (0,0)\| < \delta = \varepsilon,$$

de modo que se satisface la definición 2.13 y concluimos que

$$\begin{aligned} \lim_{(x,y)\to(0,0)} g(x,y) &= \lim_{(x,y)\to(0,0)} \frac{x^2 y}{x^2 + y^2} \\ &= 0. \end{aligned}$$

Por la proposición 2.13 tenemos entonces que

$$\begin{aligned}\lim_{(x,y)\to(0,0)} h(x,y) &= \lim_{(x,y)\to(0,0)} (f\circ g)(x,y)\\ &= \lim_{(x,y)\to(0,0)} \cos\left(\frac{x^2y}{x^2+y^2}\right)\\ &= \cos\left(\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^2+y^2}\right)\\ &= \cos(0)\\ &= 1.\end{aligned}$$

Otra pregunta que resulta interesante con relación a la proposición 2.13 es si las hipótesis sobre las funciones f y g son "intercambiables". Es decir, si ahora suponemos que g es continua en $\hat{x}_0 \in A$ y f tiene límite en $\hat{l} = g(\hat{x}_0) \in D'$, ¿se cumple que la función $f \circ g$ tiene límite en $\hat{x}_0$?

La respuesta a esta pregunta es negativa y el ejemplo 2.9 de la sección anterior nos sirve como contraejemplo. Sin embargo, tomando en cuenta las características de ese ejemplo, podemos agregar una hipótesis que nos permita formular una proposición análoga.

Proposición 2.14 *Sean $g : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $f : D \subset \mathbb{R}^m \to \mathbb{R}^k$, $\hat{x}_0 \in \mathbb{R}^n$ y $\hat{l} \in \mathbb{R}^k$ tales que $\hat{x}_0 \in A$ y $\hat{y}_0 = g(\hat{x}_0) \in D'$. Si g es continua en $\hat{x}_0$, $\lim_{\hat{y}\to\hat{y}_0} f(\hat{y}) = \hat{l}$ y existe $r > 0$ tal que $g(\hat{x}) \neq \hat{y}_0$ para toda $\hat{x} \in B_r(\hat{x}_0)\backslash\{\hat{x}_0\}$, entonces $f \circ g$ tiene límite en $\hat{x}_0$ y además*

$$\lim_{\hat{x}\to\hat{x}_0} (f\circ g)(\hat{x}) = \hat{l}.$$

La prueba de esta proposición es muy parecida a la de la proposición 2.13 y por lo mismo se deja al lector. Por otra parte, aún cuando la conclusión de esta proposición asegura la existencia de un límite, se suele usar con más frecuencia para demostrar que una función no tiene límite, como se muestra en el siguiente

Ejemplo 2.12 *Sea*

$$f(x,y) = \frac{x^2y}{x^4+y^2}.$$

Mostraremos que esta función no tiene límite en el punto $(0,0)$.

Lo que haremos será "aproximarnos" al punto $(0,0)$ *a través de diferentes "curvas", las cuales definiremos por medio de funciones de* $\mathbb{R}$ *en* $\mathbb{R}^2$ *(funciones que serán el tema principal del capítulo* 3*), y que tienen la particularidad de "pasar" por el punto que nos interesa.*

Primero consideraremos una función de la forma $g(t) = (t, mt)$, *con* $m \neq 0$ *arbitrario. Nótese que esta función claramente es continua para cualquier* $t \in \mathbb{R}$ *(proposición 2.11) y su imagen es una recta (de pendiente* m*) que "pasa" por el origen* $(0,0)$ *cuando* $t = 0$. *Si ahora realizamos la composición* $f \circ g$ *tenemos que*

$$\begin{aligned}(f \circ g)(t) &= \frac{t^2(mt)}{(t^2)^2 + (mt)^2} \\ &= \frac{mt}{t^2 + m^2}\end{aligned}$$

para toda $t \in \mathbb{R}$, *y obtenemos que*

$$\lim_{t \to 0}(f \circ g)(t) = 0.$$

Si además observamos que $f(t,0) = 0 = f(0,t)$ *para toda* $t \in \mathbb{R}$, *podemos decir que* f *se "aproxima" a* 0 *cuando nos "aproximamos" al* $(0,0)$ *a través de cualquier recta.*

Ahora consideremos la función $h(t) = \left(t, t^2\right)$. *En este caso la función es otra vez claramente continua para cualquier* $t \in \mathbb{R}$, *"describe" (o "recorre") la parábola* $y = x^2$ *y nuevamente en* $t = 0$ *"pasa" por el origen* $(0,0)$, *lo que significa que en este caso nos "aproximamos" al* $(0,0)$ *a través de esta curva. Si consideramos nuevamente la composición* $f \circ h$, *tenemos que*

$$\begin{aligned}(f \circ h)(t) &= \frac{t^2 t^2}{(t^2)^2 + (t^2)^2} \\ &= \frac{1}{2}\end{aligned}$$

para toda $t \in \mathbb{R}$, $t \neq 0$, *de modo que ahora tenemos que*

$$\lim_{t \to 0}(f \circ h)(t) = \frac{1}{2}.$$

Con base en estos dos resultados y en la proposición 2.14 podemos concluir que nuestra función f no tiene límite en el $(0,0)$.

Para terminar esta subsección (y antes de pasar a los teoremas "fuertes" de continuidad), definiremos lo que significa que una función sea continua en un subconjunto de su dominio (definición que al lector le resultará del todo "natural"), y probaremos una interesante y útil caracterización de esta propiedad.

Definición 2.15 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $B \subset A$. Decimos que f es continua en (o sobre) B si f es continua en cada punto de B, es decir, si f es continua para cada $\hat{x} \in B$.*

La siguiente caracterización de la continuidad de un función f sobre su dominio A será de gran utilidad y podemos motivarla a partir de la contención que escribimos al final de la definición 2.14. En efecto, recordemos que si f es continua en un punto $\hat{x} \in A$, podemos asegurar que para cada $\varepsilon > 0$ existe $\delta > 0$ tal que

$$f(B_\delta(\hat{x}) \cap A) \subset B_\varepsilon(f(\hat{x}))$$

y, por lo tanto, que

$$B_\delta(\hat{x}) \cap A \subset f^{-1}\left(B_\varepsilon(f(\hat{x}))\right).$$

De esta forma, si $V \subset \mathbb{R}^m$ es un conjunto abierto y $\hat{x} \in f^{-1}(V)$, se tiene que $f(\hat{x}) \in V$ de modo que, como V es abierto, existe $\varepsilon_{\hat{x}} > 0$ tal que $B_\varepsilon(f(\hat{x})) \subset V$, y como f es continua en $\hat{x}$ (pues es continua en A), existe $\delta_{\hat{x}} > 0$ tal que

$$B_{\delta_{\hat{x}}}(\hat{x}) \cap A \subset f^{-1}\left(B_{\varepsilon_{\hat{x}}}(f(\hat{x}))\right).$$

Si hacemos lo anterior para cada $\hat{x} \in f^{-1}(V)$, por las propiedades de la imagen inversa dadas en la proposición 2.1 tendremos que

$$\begin{aligned}\left(\bigcup_{\hat{x}\in f^{-1}(V)} B_{\delta_{\hat{x}}}(\hat{x})\right)\cap A &= \bigcup_{\hat{x}\in f^{-1}(V)} (B_{\delta_{\hat{x}}}(\hat{x})\cap A)\\ &\subset \bigcup_{\hat{x}\in f^{-1}(V)} f^{-1}\left(B_{\varepsilon_{\hat{x}}}(f(\hat{x}))\right)\\ &\subset f^{-1}(V)\\ &\subset \left(\bigcup_{\hat{x}\in f^{-1}(V)} B_{\delta_{\hat{x}}}(\hat{x})\right)\cap A,\end{aligned}$$

es decir, que

$$f^{-1}(V) = \left(\bigcup_{\hat{x}\in f^{-1}(V)} B_{\delta_{\hat{x}}}(\hat{x})\right)\cap A.$$

Si ahora recordamos que las bolas (o vecindades) son conjuntos abiertos, por el problema 21 del capítulo 1 tenemos que

$$U = \bigcup_{\hat{x}\in f^{-1}(V)} B_{\delta_{\hat{x}}}(\hat{x})$$

es un conjunto abierto, de tal forma que lo que hemos probado es que $f^{-1}(V) = U\cap A$. Es decir, que la imagen inversa $f^{-1}(V)$ se puede ver como la intersección de un conjunto abierto U de $\mathbb{R}^n$ con el dominio A de f.

Lo que formularemos y probaremos en la siguiente proposición es que esta propiedad no sólo es una consecuencia necesaria de la continuidad de f en A, sino que también es una condición suficiente para ella.

Proposición 2.15 *Sea $f : A\subset\mathbb{R}^n\to\mathbb{R}^m$. La función f es continua en A si y sólo si para todo conjunto abierto $V\subset\mathbb{R}^m$ existe un conjunto abierto $U\subset\mathbb{R}^n$ tal que $f^{-1}(V) = A\cap U$.*

Demostración.
$\Rightarrow]$
La prueba de esta implicación la hicimos en los párrafos anteriores, así que no la repetiremos. Sólo restaría agregar que si $f^{-1}(V) = \varnothing$, entonces tomamos $U = \varnothing$.
$\Leftarrow]$
Sea $\hat{x} \in A$ y $\varepsilon > 0$. Como la bola $B_\varepsilon(f(\hat{x})) \subset \mathbb{R}^m$ es un conjunto abierto, por hipótesis existe $U \subset \mathbb{R}^n$ abierto tal que $f^{-1}(B_\varepsilon(f(\hat{x}))) = U \cap A$. Ahora, dado que $\hat{x} \in f^{-1}(B_\varepsilon(f(\hat{x}))) = U \cap A$ y U es abierto, existe $\delta > 0$ tal que $B_\delta(\hat{x}) \subset U$, de modo que

$$B_\delta(\hat{x}) \cap A \subset U \cap A = f^{-1}(B_\varepsilon(f(\hat{x})))$$

y por lo tanto $f(B_\delta(\hat{x}) \cap A) \subset f(f^{-1}(B_\varepsilon(f(\hat{x})))) \subset B_\varepsilon(f(\hat{x}))$, lo que prueba que f es continua en $\hat{x}$. ■

2.2.4. Teoremas "fuertes" de continuidad

Seguramente el lector estará de acuerdo en que tratándose de funciones continuas de $\mathbb{R}$ en $\mathbb{R}$ dos son los resultados más importantes relacionados con ellas: el *Teorema del Valor Intermedio*, y el *Teorema del valor Máximo (y el valor Mínimo)*, que asegura que toda función continua sobre un intervalo de la forma $[a, b]$ (un subconjunto cerrado y acotado (además de conexo) de $\mathbb{R}$) siempre alcanza su valor máximo y su valor mínimo.

Sin restarles importancia a estos teoremas, lo que es más relevante aún, son los dos resultados que podemos probar o reformular a partir de ellos (o de los argumentos usados en su prueba). El primero de ellos, que se prueba a partir del Teorema del Valor Intermedio, es aquel que asegura que si se tiene función continua sobre un intervalo $I \subset \mathbb{R}$ (totalmente arbitrario, sin importar si I es abierto, cerrado, acotado o no acotado), entonces su imagen $f(I) \subset \mathbb{R}$ también es un intervalo.

Si observamos que los intervalos son los únicos subconjuntos conexos de $\mathbb{R}$ (problema 40, capítulo 1), el resultado que acabamos de mencionar se

podría reformular de la siguiente manera: si f es continua sobre el conjunto $I \subset \mathbb{R}$, e I es conexo, entonces $f(I) \subset \mathbb{R}$ es conexo.

El segundo resultado importante que se puede probar usando los mismos argumentos que se usan para la prueba del Teorema del valor Máximo (y el valor Mínimo) es aquel que asegura que si $A \subset \mathbb{R}$ es un conjunto cerrado y acotado (aún cuando este conjunto A no sea de la forma $[a, b]$), entonces $f(A) \subset \mathbb{R}$ también es un conjunto cerrado y acotado.

Lo que vamos a hacer en esta subsección es mostrar que los resultados que acabamos de mencionar (para funciones de $\mathbb{R}$ en $\mathbb{R}$) se siguen cumpliendo para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$.

Teorema 2.3 *Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ continua en A y $B \subset A$. Si B es conexo, entonces $f(B)$ es conexo.*

Demostración. En esta prueba, además de hacerla por el método de la contrapuesta, usaremos la equivalencia para conjuntos disconexos probada en la proposición 1.14 del capítulo 1. Supongamos entonces que $f(B)$ no es conexo; por la proposición mencionada, sabemos que existen $\tilde{U}, \tilde{V} \subset \mathbb{R}^m$ conjuntos abiertos tales que:

(a) $f(B) \subset \tilde{U} \cup \tilde{V}$,

(b) $f(B) \cap \tilde{U} \neq \varnothing$ y $f(B) \cap \tilde{V} \neq \varnothing$, y

(c) $f(B) \cap \tilde{U} \cap \tilde{V} = \varnothing$.

Por la proposición 2.15 sabemos que existen $U, V \subset \mathbb{R}^n$ conjuntos abiertos tales que $f^{-1}(\tilde{U}) = A \cap U$ y $f^{-1}(\tilde{V}) = A \cap V$. Afirmamos que U y V satisfacen que:

(1) $B \subset U \cup V$,

(2) $B \cap U \neq \varnothing$ y $B \cap V \neq \varnothing$, y

(3) $B \cap U \cap V = \varnothing$.

Para probar lo anterior usaremos varias de las propiedades de la imagen inversa y la imagen directa formuladas en la proposición 2.1.

Por el inciso (a) de nuestra suposición, sabemos que

$$\begin{aligned} B &\subset f^{-1}(f(B)) \\ &\subset f^{-1}(\tilde{U} \cup \tilde{V}) \\ &= f^{-1}(\tilde{U}) \cup f^{-1}(\tilde{V}) \\ &= (A \cap U) \cup (A \cap V) \\ &= A \cap (U \cup V) \end{aligned}$$

y por lo tanto $B \subset U \cup V$, con lo cual probamos el inciso (1).

Ahora, por el inciso (b), como $f(B) \cap \tilde{U} \neq \varnothing$, entonces existe $\hat{x} \in B$ tal que $f(\hat{x}) \in \tilde{U}$. Por tanto, $\hat{x} \in f^{-1}(\tilde{U}) = A \cap U$, de modo que $\hat{x} \in B \cap U$, es decir, $B \cap U \neq \varnothing$. Análogamente se prueba que $B \cap V \neq \varnothing$, con lo cual tenemos probado el inciso (2).

Finalmente, dado que $B \subset A$ y $B \subset f^{-1}(f(B))$, tenemos que

$$\begin{aligned} B \cap U \cap V &= B \cap (U \cap A) \cap (V \cap A) \\ &= B \cap f^{-1}(\tilde{U}) \cap f^{-1}(\tilde{V}) \\ &\subset f^{-1}(f(B)) \cap f^{-1}(\tilde{U}) \cap f^{-1}(\tilde{V}) \\ &= f^{-1}(f(B) \cap \tilde{U} \cap \tilde{V}) \\ &= f^{-1}(\varnothing) \\ &= \varnothing, \end{aligned}$$

de donde $B \cap U \cap V = \varnothing$, con lo cual probamos el inciso (3).

Como el lector habrá notado, los incisos (1), (2) y (3), por la misma proposición 2.1, implican que B es disconexo, conclusión con la cual terminamos la prueba. ■

Con base en el teorema anterior, se establece una interesante versión del teorema del valor intermedio para funciones de $\mathbb{R}^n$ en $\mathbb{R}$, resultado que dejamos plasmado en el siguiente corolario y cuya prueba se deja al lector.

Corolario 2.2 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}$ continua en A, $B \subset A$ conexo y $\hat{x}_1, \hat{x}_2 \in B$ tales que $f(\hat{x}_1) < f(\hat{x}_2)$. Si $c \in \mathbb{R}$ es tal que $f(\hat{x}_1) < c < f(\hat{x}_2)$, entonces existe $\hat{x} \in B$ tal que $f(\hat{x}) = c$.*

Como mencionamos al inicio de esta subsección, otro tipo de conjuntos que "preservan" sus características bajo funciones continuas son los conjuntos cerrados y acotados. Este resultado tiene consecuencias muy importantes, en particular las relacionadas con la existencia de valores máximos y mínimos de funciones de $\mathbb{R}^n$ en $\mathbb{R}$, tema que trataremos ampliamente en el capítulo 4.

Teorema 2.4 *Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ continua en A y $B \subset A$. Si B es cerrado y acotado, entonces $f(B)$ es cerrado y acotado.*

Demostración. En la prueba de este teorema jugarán un papel muy importante varios de los resultados que probamos para sucesiones.

Primero probaremos que $f(B)$ está acotado y usaremos nuevamente el método de la contrapuesta. Supongamos entonces que $f(B)$ no está acotado. Bajo este supuesto, para cada $k \in \mathbb{N}$ existe $\hat{x}_k \in B$ tal que $\|f(\hat{x}_k)\| > k$.

Dado que B está acotado, entonces $\{\hat{x}_k\}$ es una sucesión acotada, de modo que por el teorema 2.1 existe una subsucesión $\{\hat{x}_{k_l}\}$ que converge. Si $\hat{x}_0 \in \mathbb{R}^n$ es tal que $\{\hat{x}_{k_l}\} \to \hat{x}_0$ (cuando $l \to \infty$), por el inciso (a) del problema 29 se tiene que $\hat{x}_0 \in \bar{B}$ y como B es un conjunto cerrado, entonces $B = \bar{B}$ y por lo tanto $\hat{x}_0 \in B$.

Ahora, como f es continua en $\hat{x}_0$, se debe tener que $\{f(\hat{x}_{k_l})\} \to f(\hat{x}_0)$ (cuando $l \to \infty$), lo cual contradice el hecho de que $\|f(\hat{x}_{k_l})\| > k_l$ para toda $l \in \mathbb{N}$ y que $\{k_l\}$ es una sucesión creciente de naturales.

Para probar que $f(B)$ es un conjunto cerrado mostraremos que $(f(B))' \subset f(B)$. Si $(f(B))' = \varnothing$ la contención es inmediata. Supongamos entonces que $\hat{y}_0 \in (f(B))'$; por el inciso (b) del mismo problema 29, sabemos que existe una sucesión $\{\hat{y}_k\}$ contenida en $f(B)$ tal que $\{\hat{y}_k\} \to \hat{y}_0$, de tal forma que si $\hat{x}_k \in B$ es tal que $f(\hat{x}_k) = \hat{y}_k$, entonces $\{\hat{x}_k\}$, por estar contenida en B, es una sucesión acotada.

Nuevamente, por el teorema 2.1 existe una subsucesión $\{\hat{x}_{k_l}\}$ que converge. Si $\hat{x}_0 \in \mathbb{R}^n$ es tal que $\{\hat{x}_{k_l}\} \to \hat{x}_0$ (cuando $l \to \infty$), por el mismo

argumento usado en la primera parte de esta prueba se tiene que $\hat{x}_0 \in B$ y como f es continua en $\hat{x}_0$, entonces $\{f(\hat{x}_{k_l}) = \hat{y}_{k_l}\} \to f(\hat{x}_0)$ (cuando $l \to \infty$).

Por otra parte, dado que $\{\hat{y}_{k_l}\}$ es una subsucesión de $\{\hat{y}_k\}$, entonces $\{\hat{y}_{k_l}\} \to \hat{y}_0$ y como el punto de convergencia de una sucesión es único, se tiene que $\hat{y}_0 = f(\hat{x}_0)$, lo que prueba que $\hat{y}_0 \in f(B)$. ■

Como anunciamos anteriormente, con base en este teorema podemos probar un resultado muy importante que nos proporciona condiciones suficientes para que una función continua de $\mathbb{R}^n$ en $\mathbb{R}$ "alcance" valores máximos y mínimos sobre un conjunto.

Este resultado lo dejamos formulado en el siguiente corolario y su prueba (¡nuevamente!) se deja al lector.

Corolario 2.3 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}$ continua en A y $B \subset A$. Si $B \neq \varnothing$ es cerrado y acotado, entonces existen $\hat{x}_1, \hat{x}_2 \in B$ tales que $f(\hat{x}_1) \leq f(\hat{x}) \leq f(\hat{x}_2)$ para toda $\hat{x} \in B$. Es decir, f alcanza un valor máximo y un valor mínimo sobre B.*

2.3. Continuidad uniforme

Concluimos este capítulo definiendo el concepto de continuidad uniforme para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$. La forma explícita de esta definición es totalmente equivalente a la de las funciones de $\mathbb{R}$ en $\mathbb{R}$ y expresa la misma propiedad de una función f sobre un conjunto B: a saber, que *para cada cantidad $\varepsilon > 0$ existe una cantidad $\delta > 0$ tal que $f(B_\delta(\hat{x}) \cap B) \subset B_\varepsilon(f(\hat{x}))$ para toda $\hat{x} \in B$*. Es decir, dada una $\varepsilon > 0$ se puede encontrar una $\delta > 0$ que "*sirve*" *para cualquier $\hat{x} \in B$* (en el sentido de que $f(B_\delta(\hat{x}) \cap B) \subset B_\varepsilon(f(\hat{x}))$), propiedad que sin duda "*dice*" que, además de que f es continua en cada $\hat{x} \in B$, esta continuidad tiene cierta cualidad de "*uniformidad*" sobre este conjunto.

Definición 2.16 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $B \subset A$. Decimos que f es uniformemente continua sobre B si para cada $\varepsilon > 0$ existe $\delta > 0$ con la propiedad de que: si $\hat{x}, \hat{y} \in B$ son tales que $\|\hat{x} - \hat{y}\| < \delta$, entonces $\|f(\hat{x}) - f(\hat{y})\| < \varepsilon$. Es decir, $f(B_\delta(\hat{x}) \cap B) \subset B_\varepsilon(f(\hat{x}))$ para toda $\hat{x} \in B$.*

Como ha sucedido con los conceptos de límite y de continuidad para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$, que se pueden caracterizar en términos de los conceptos correspondientes de sus funciones coordenadas, para el caso de la continuidad uniforme sucede lo mismo, propiedad que vamos a dejar expresada en la siguiente proposición y cuya prueba ... (¡adivine el lector!).

Proposición 2.16 *Sean $f = (f_1, \ldots, f_m) : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $B \subset A$. La función f es uniformemente continua sobre B si y sólo si cada función f_i es uniformemente continua sobre B, para $i \in \{1, \ldots, m\}$.*

Como mencionamos en el comentario previo a la definición 2.16, toda función que es uniformemente continua sobre un conjunto A también es continua sobre el mismo conjunto, hecho que dejamos plasmado en la siguiente proposición y cuya veracidad es tan inmediatamente clara, que omitiremos su prueba.

Proposición 2.17 *Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$. Si f es uniformemente continua sobre A, entonces f es continua en A.*

Como seguramente el lector tendrá presente para el caso de las funciones de $\mathbb{R}$ en $\mathbb{R}$, el recíproco de la proposición anterior es falso, situación que se repite para las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$, como lo mostraremos en el siguiente

Ejemplo 2.13 *Sea*

$$\begin{aligned} f(x,y) &= \frac{1}{\sqrt{x^2+y^2}} \\ &= \frac{1}{\|(x,y)\|} \end{aligned}$$

con $(x,y) \in A = \mathbb{R}^2 \setminus \{(0,0)\}$.

Mostraremos en general que, si $B \subset A$ es tal que $(0,0) \in B'$, entonces f no es uniformemente continua en B.

Este ejemplo, además de mostrar que el recíproco de la proposición 2.17 no se satisface, tendrá el mérito de exhibir un método para probar que una función no es uniformemente continua sobre un conjunto.

De acuerdo con la definición 2.16, para probar que f no es uniformemente continua sobre el conjunto B es necesario mostrar que existe una cantidad específica $\varepsilon > 0$ tal que para cualquier $\delta > 0$, siempre se pueden encontrar puntos $\hat{x}_1, \hat{x}_2 \in B$ tales que $\|\hat{x}_1 - \hat{x}_2\| < \delta$ y sin embargo $|f(\hat{x}_1) - f(\hat{x}_2)| \geq \varepsilon$.

Tomemos $\varepsilon = 1$ y sea $\delta > 0$ arbitrario. Como $(0,0) \in B'$ existe $\hat{x}_1 \in B$ tal que

$$0 < \|\hat{x}_1\| = \|\hat{x}_1 - (0,0)\| < \delta/2$$

y por la misma razón, existe $\hat{x}_2 \in B$ tal que

$$\begin{aligned} \|\hat{x}_2\| &= \|\hat{x}_2 - (0,0)\| \\ &< \frac{\|\hat{x}_1\|}{\|\hat{x}_1\| + 1} \\ &< \|\hat{x}_1\|, \end{aligned}$$

de tal forma que $0 < \|\hat{x}_2\| < \|\hat{x}_1\|$. Por lo tanto, $\hat{x}_1, \hat{x}_2 \in B$ son tales que

$$\begin{aligned} \|\hat{x}_1 - \hat{x}_2\| &\leq \|\hat{x}_1\| + \|\hat{x}_2\| \\ &< 2\,\|\hat{x}_1\| \\ &< \delta \end{aligned}$$

y además tienen la propiedad de que

$$\begin{aligned} |f(\hat{x}_1) - f(\hat{x}_2)| &= \left| \frac{1}{\|\hat{x}_1\|} - \frac{1}{\|\hat{x}_2\|} \right| \\ &= \frac{1}{\|\hat{x}_2\|} - \frac{1}{\|\hat{x}_1\|} \\ &> \frac{\|\hat{x}_1\| + 1}{\|\hat{x}_1\|} - \frac{1}{\|\hat{x}_1\|} \\ &= 1, \end{aligned}$$

que es lo que deseábamos probar.

El ejemplo anterior también tiene la virtud de mostrar el tipo de características que deben de tener una función f y un conjunto B, si se desea concluir que f no es uniformemente continua sobre B.

En efecto, para concluir lo anterior basta con mostrar que en el conjunto B existen puntos tan cercanos como se quiera, tales que sus imágenes bajo la función f se "alejan" entre ellas (o permanecen a una distacia mayor que una constante). Y esto es justo lo que sucede con aquellos puntos que están "cerca" del $(0,0)$ en el ejemplo anterior, en donde, si evaluamos a f en puntos cada vez más cercanos a este punto, los valores que se obtienen se hacen cada vez más grandes (y están más "alejados" entre ellos).

Por otra parte, en el problema 62 se pide al lector probar que la misma función f del ejemplo anterior sí es uniformemente continua sobre cualquier subconjunto B de su dominio, siempre y cuando se satisfaga que el $(0,0)$ no sea punto de acumulación de B.

Con esta prueba se mostrará que la continuidad uniforme es un concepto que depende no sólo de la función que estemos tratando, sino también del conjunto sobre el cual la estemos considerando. La misma función puede ser uniformemente continua sobre algunos conjuntos, y no serlo sobre otros. No tiene sentido afirmar que una función es (o no es) uniformemente continua, si no se menciona un conjunto sobre el cual se pretende que lo sea (o que no lo sea).

Regresando al recíproco de la proposición 2.17, dado que a la función involucrada no le podemos pedir algo más que ser continua, la opción es pedirle algunas características al conjunto sobre el cuál queremos que ésta sea uniformemente continua. Una manera de dar con estas características (que en matemáticas suele dar sus frutos), consiste en intentar hacer la prueba y sobre la marcha ubicar qué es lo que se va "necesitando". ¡Manos a la obra!

Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, con f continua sobre A, y una cantidad $\varepsilon > 0$ arbitraria. La continuidad de f en cada punto de $\hat{x} \in A$ nos asegura que para cada uno de ellos existe $\delta_{\hat{x}} > 0$ tal que $f(B_{\delta_{\hat{x}}}(\hat{x}) \cap A) \subset B_{\varepsilon}(f(\hat{x}))$.

Sin duda que nuestro primer impulso para encontrar una sola $\delta > 0$ para la cual la contención anterior se cumpla para toda $\hat{x} \in A$, sería tomar "la mínima" (o el ínfimo, para ser más precisos) de entre todas las $\delta_{\hat{x}}$, pero a

estas alturas ya sabemos que "esa mínima" (o el ínfimo) de todas ellas no tiene por qué ser mayor que cero. Ante esta situación, vale la pena observar lo siguiente: el conjunto A siempre está contenido (o queda "cubierto") por la unión de todas las vecindades $B_{\delta_{\hat{x}}}(\hat{x})$, es decir

$$A \subset \bigcup_{\hat{x}\in A} B_{\delta_{\hat{x}}}(\hat{x}).$$

Más aún, podemos tomar la mitad de cada uno de los radios $\delta_{\hat{x}}$ (o la tercera parte, o la cuarta parte, ¡o cualquier otra cantidad menor!) y se sigue cumpliendo que

$$A \subset \bigcup_{\hat{x}\in A} B_{\frac{\delta_{\hat{x}}}{2}}(\hat{x}).$$

Si de entre todas estas vecindades (o bolas) se puede encontrar un número finito de ellas que sigan teniendo la propiedad de que "cubren" a A, es decir que existen $\hat{x}_1, \ldots, \hat{x}_k \in A$ tales que

$$A \subset B_{\frac{\delta_{\hat{x}_1}}{2}}(\hat{x}_1) \cup \cdots \cup B_{\frac{\delta_{\hat{x}_k}}{2}}(\hat{x}_k),$$

entonces podemos elegir $\delta = \min\{\delta_{\hat{x}_1}/2, \ldots, \delta_{\hat{x}_k}/2\}$ (la cual sí será mayor que 0, puesto que ahora sí estamos tomando el mínimo de un conjunto finito), cantidad para la cual se va a satisfacer el siguiente hecho: si $\hat{x}, \hat{y} \in A$ son tales que $\|\hat{x} - \hat{y}\| < \delta$, entonces existe $j \in \{1, \ldots, k\}$ tal que

$$\hat{x}, \hat{y} \in B_{\delta_{\hat{x}_j}}(\hat{x}_j),$$

es decir, que $\hat{x}$ y $\hat{y}$ pertenecen a la misma vecindad.

La afirmación anterior (de la que más adelante daremos su prueba) resulta ser muy importante, puesto que ahora, dado que $f(B_{\delta_{\hat{x}_j}}(\hat{x}_j) \cap A) \subset B_\varepsilon(f(\hat{x}_j))$, por una sencilla aplicación de la desigualdad del triángulo, tendremos que

$$\begin{aligned}
\|f(\hat{x}) - f(\hat{y})\| &= \|(f(\hat{x}) - f(\hat{x}_j)) - (f(\hat{x}_j) - f(\hat{y}))\| \\
&\leq \|f(\hat{x}) - f(\hat{x}_j)\| + \|f(\hat{x}_j) - f(\hat{y})\| \\
&< \varepsilon + \varepsilon \\
&= 2\varepsilon
\end{aligned}$$

y por lo tanto, ya que ε fue una cantidad arbitraria, podremos concluir que ¡f es uniformemente continua!

Como seguramente el lector habrá notado, la hipótesis que resulta determinante en la argumentación que acabamos de dar está en el supuesto de que, para cada familia (o conjunto) de vecindades $\left\{B_{\delta_{\hat{x}}/2}(\hat{x}) \mid \hat{x} \in A\right\}$ que "cubren" al conjunto A (familia de vencindades que depende de cada cantidad $\varepsilon > 0$ que se tome), existe una subfamilia *finita* que sigue teniendo la propiedad de "cubrir" a A. Los conjuntos que tienen esta propiedad (o una equivalente, en donde las vecindades se sustituyen por conjuntos abiertos arbitrarios), reciben el nombre de *conjuntos compactos*, los cuales introducimos en la siguiente

Definición 2.17 *Sean $K \subset \mathbb{R}^n$ y $\{U_\alpha \subset \mathbb{R}^n \mid \alpha \in I\}$ una familia de subconjuntos abiertos de $\mathbb{R}^n$ indexada por un conjunto I. Decimos que:*

1. *la familia de subconjuntos $\mathcal{U} = \{U_\alpha \subset \mathbb{R}^n \mid \alpha \in I\}$ es una cubierta (abierta) de K si*
$$K \subset \bigcup_{\alpha \in I} U_\alpha$$
2. *el conjunto K es un conjunto compacto si toda cubierta (abierta) $\mathcal{U} = \{U_\alpha \subset \mathbb{R}^n \mid \alpha \in I\}$ de K tiene una subcubierta finita, es decir, si existen $\alpha_1, \ldots, \alpha_k \in I$ tales que*
$$K \subset U_{\alpha_1} \cup \cdots \cup U_{\alpha_k}.$$

Con base en esta definición, ya estamos en condiciones de formular un resultado que da respuesta a la pregunta que hicimos sobre el recíproco de la proposición 2.17, el cual dejaremos plasmado en el siguiente

Teorema 2.5 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $K \subset A$. Si f es continua sobre K y K es un conjunto compacto, entonces f es uniformemente continua sobre K.*

Demostración. Sea $\varepsilon > 0$. Como f es continua para cada $\hat{x} \in K$, sabemos que existe $\delta_{\hat{x}} > 0$ tal que $f(B_{\delta_{\hat{x}}}(\hat{x}) \cap K) \subset B_{\varepsilon/2}(f(\hat{x}))$. Dado que la familia de

vecindades $\mathcal{U} = \{B_{\delta_{\hat{x}}/2}(\hat{x}) \mid \hat{x} \in K\}$ es una cubierta abierta de K, sabemos que existen $\hat{x}_1, \ldots, \hat{x}_k \in K$ tales que

$$K \subset B_{\frac{\delta_{\hat{x}_1}}{2}}(\hat{x}_1) \cup \cdots \cup B_{\frac{\delta_{\hat{x}_k}}{2}}(\hat{x}_k). \tag{2.9}$$

Tomamos

$$\delta = \text{mín}\left\{\frac{\delta_{\hat{x}_1}}{2}, \ldots, \frac{\delta_{\hat{x}_k}}{2}\right\} > 0$$

y sean $\hat{x}, \hat{y} \in K$ tales que $\|\hat{x} - \hat{y}\| < \delta$. Por la contención 2.9 sabemos que existe $j \in \{1, \ldots, k\}$ tal que

$$\hat{x} \in B_{\frac{\delta_{\hat{x}_j}}{2}}(\hat{x}_j),$$

y como $\|\hat{x} - \hat{y}\| < \delta \leq \delta_{\hat{x}_j}/2$ se tiene que

$$\begin{aligned} \|\hat{y} - \hat{x}_j\| &= \|(\hat{y} - \hat{x}) + (\hat{x} - \hat{x}_j)\| \\ &\leq \|\hat{y} - \hat{x}\| + \|\hat{x} - \hat{x}_j\| \\ &< \delta + \frac{\delta_{\hat{x}_j}}{2} \\ &\leq \delta_{\hat{x}_j}, \end{aligned}$$

de donde $\hat{y} \in B_{\delta_{\hat{x}_j}}(\hat{x}_j)$. Así, como $\hat{x}, \hat{y} \in B_{\delta_{\hat{x}_j}}(\hat{x}_j) \cap K$ concluimos que

$$\begin{aligned} \|f(\hat{x}) - f(\hat{y})\| &= \|(f(\hat{x}) - f(\hat{x}_j)) - (f(\hat{x}_j) - f(\hat{y}))\| \\ &\leq \|f(\hat{x}) - f(\hat{x}_j)\| + \|f(\hat{x}_j) - f(\hat{y})\| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \end{aligned}$$

y por lo tanto, que f es uniformemente continua sobre K. ■

No podemos ignorar que al lector le parecerá un tanto "mágica" la forma en que hemos introducido el concepto de conjunto compacto, razón por la cual dedicaremos los últimos resultados de esta sección (¡y de este capítulo!)

a establecer una caracterización (en $\mathbb{R}^n$) muy importante de este tipo de conjuntos.

En efecto, existe un conocido y muy importante teorema que da una caracterización de los conjuntos compactos, el teorema de Heine-Borel[3], el cual establece que en $\mathbb{R}^n$, los conjuntos compactos no son más que nuestros muy conocidos conjuntos ¡cerrados y acotados!

Teorema 2.6 (de Heine-Borel) *Sea $K \subset \mathbb{R}^n$. K es un conjunto compacto si y sólo si K es un conjunto cerrado y acotado.*

Demostración.
$\Rightarrow]$
Primero mostraremos que K es un conjunto acotado. Para ello, consideremos la familia $\mathcal{U} = \{B_k(\hat{0}) \mid k \in \mathbb{N}\}$ de subconjuntos abiertos de $\mathbb{R}^n$. Dado que

$$\bigcup_{k\in\mathbb{N}} B_k(\hat{0}) = \mathbb{R}^n,$$

entonces $\mathcal{U}$ es una cubierta abierta de K, de modo que existen $k_1, \ldots, k_l \in \mathbb{N}$ tales que

$$K \subset B_{k_1}(\hat{0}) \cup \cdots \cup B_{k_l}(\hat{0}),$$

de tal forma que, si $N = \text{máx}\{k_1, \ldots, k_l\}$ se tiene que

$$\begin{aligned} K &\subset B_{k_1}(\hat{0}) \cup \cdots \cup B_{k_l}(\hat{0}) \\ &= B_N(\hat{0}), \end{aligned}$$

de donde concluimos que K está acotado.

[3]Heinrich Eduard Heine (16 de marzo de 1821, Berlin–21 de octubre de 1881, Halle). Heine es conocido por sus resultados sobre funciones especiales y en análisis real. En particular, escribió un importante tratado sobre funciones armónicas esféricas y funciones de Legendre. También investigó series hipergeométricas básicas e introdujo la fórmula Mehler–Heine.

Félix Édouard Justin Émile Borel (7 de enero de 1871–3 de Febrero 1956) fue un matemático y político francés. Como matemático, es conocido por su trabajo fundacional en las áreas de teoría de la medida y probabilidad. (fuente: Wikipedia).

Para probar que K es cerrado, mostraremos que $K^c := \mathbb{R}^n \setminus K$ es un conjunto abierto. Sea $\hat{x} \in \mathbb{R}^n \setminus K$ y para cada $k \in \mathbb{N}$ hacemos $U_k = \mathbb{R}^n \setminus \overline{B_{1/k}(\hat{x})}$; dado que $\overline{B_{1/k}(\hat{x})} = \{\hat{y} \in \mathbb{R}^n \mid \|\hat{y} - \hat{x}\| \leq 1/k\}$ es un conjunto cerrado, U_k es un conjunto abierto, y como

$$\bigcap_{k\in\mathbb{N}} \overline{B_{1/k}(\hat{x})} = \{\hat{x}\},$$

se tiene que

$$\begin{aligned} K &\subset \mathbb{R}^n \setminus \{\hat{x}\} \\ &= \mathbb{R}^n \setminus \bigcap_{k\in\mathbb{N}} \overline{B_{1/k}(\hat{x})} \\ &= \bigcup_{k\in\mathbb{N}} \left(\mathbb{R}^n \setminus \overline{B_{1/k}(\hat{x})}\right) \\ &= \bigcup_{k\in\mathbb{N}} U_k \end{aligned}$$

de tal forma que $\mathcal{U} = \{U_k \mid k \in \mathbb{N}\}$ es una cubierta abierta de K. Como K es compacto, existen $k_1, \ldots, k_l \in \mathbb{N}$ tales que

$$\begin{aligned} K &\subset U_{k_1} \cup \cdots \cup U_{k_l} \\ &= \left(\mathbb{R}^n \setminus \overline{B_{1/k_1}(\hat{x})}\right) \cup \cdots \cup \left(\mathbb{R}^n \setminus \overline{B_{1/k_l}(\hat{x})}\right) \\ &= \mathbb{R}^n \setminus \overline{B_{1/N}(\hat{x})} \end{aligned}$$

en donde $N = \text{máx}\{k_1, \ldots, k_l\}$. Por tanto,

$$\begin{aligned} B_{1/N}(\hat{x}) &\subset \overline{B_{1/N}(\hat{x})} \\ &\subset \mathbb{R}^n \setminus K, \end{aligned}$$

con lo que concluimos que $\mathbb{R}^n \setminus K$ es abierto y por lo tanto, que K es cerrado.
$\Leftarrow]$
Supongamos ahora que $K \subset \mathbb{R}^n$ es un conjunto cerrado y acotado.

Para esta implicación procederemos por contradicción. Supongamos entonces que K no es compacto. Bajo este supuesto, entonces debe existir una

cubierta abierta $\mathcal{U} = \{U_\alpha \mid \alpha \in I\}$ de K tal que ningún número finito de elementos de $\mathcal{U}$ sigue cubriendo a K. Ahora, por el problema 20 del capítulo 1, sabemos que cada U_α se puede expresar como la unión de bolas (o vecindades) con centro en un punto de $\mathbb{Q}^n$ y radio racional.

Dado que $\mathbb{Q}^n$ y los racionales (positivos) son conjuntos numerables, entonces la familia $\mathcal{V}$ de todas las vecindades de este tipo que son necesarias para expresar a todos los U_α como unión de éstas, también es numerable, y por lo tanto existe una sucesión $\{\hat{x}_k\}$ de puntos en $\mathbb{Q}^n$ y una sucesión $\{r_k\}$ de números racionales positivos tales que la familia de bolas $\mathcal{V} = \{B_{r_k}(\hat{x}_k) \mid k \in \mathbb{N}\}$ satisface las siguientes condiciones:

1. para cada $k \in \mathbb{N}$ existe $\alpha \in I$ tal que $B_{r_k}(\hat{x}_k) \subset U_\alpha$, y
2. $$\bigcup_{k\in\mathbb{N}} B_{r_k}(\hat{x}_k) = \bigcup_{\alpha\in I} U_\alpha.$$

Nótese que, dado que no existe un número finito de elementos de $\mathcal{U}$ que cubra a K, por el inciso 1 tampoco existe un número finito de elementos de $\mathcal{V}$ que cubra a K; y por el inciso 2, dado que $\mathcal{U}$ es una cubierta abierta de K, $\mathcal{V}$ también es una cubierta abierta de K.

Ahora, para cada $k \in \mathbb{N}$, hacemos

$$A_k = (\mathbb{R}^n \setminus (B_{r_1}(\hat{x}_1) \cup \cdots \cup B_{r_k}(\hat{x}_k))) \cap K.$$

Obsérvese que estos conjuntos A_k satisfacen las siguientes condiciones, para cada $k \in \mathbb{N}$:

a. $A_k \neq \varnothing$ (recuerde que no existe un número finito de elementos de $\mathcal{V}$ que cubra a K),

b. $A_{k+1} \subset A_k$ (pues $B_{r_1}(\hat{x}_1) \cup \cdots \cup B_{r_k}(\hat{x}_k) \subset B_{r_1}(\hat{x}_1) \cup \cdots \cup B_{r_k}(\hat{x}_k) \cup B_{r_{k+1}}(\hat{x}_{k+1})$), y

c. A_k es cerrado y acotado (es la intersección de dos conjuntos cerrados, uno de los cuales, K, está acotado)

Por tanto, por el problema 34 del capítulo 1, se tiene que

$$\begin{aligned}
\varnothing &\neq \bigcap_{k\in\mathbb{N}} A_k \\
&= \bigcap_{k\in\mathbb{N}} (\mathbb{R}^n \setminus (B_{r_1}(\hat{x}_1) \cup \cdots \cup B_{r_k}(\hat{x}_k))) \cap K \\
&= \left(\bigcap_{k\in\mathbb{N}} (\mathbb{R}^n \setminus (B_{r_1}(\hat{x}_1) \cup \cdots \cup B_{r_k}(\hat{x}_k))) \right) \cap K \\
&= \left(\mathbb{R}^n \setminus \bigcup_{k\in\mathbb{N}} (B_{r_1}(\hat{x}_1) \cup \cdots \cup B_{r_k}(\hat{x}_k)) \right) \cap K \\
&= \left(\mathbb{R}^n \setminus \bigcup_{k\in\mathbb{N}} B_{r_k}(\hat{x}_k) \right) \cap K.
\end{aligned}$$

Por otra parte, como $\mathcal{V} = \{B_{r_k}(\hat{x}_k) \mid k \in \mathbb{N}\}$ también es una cubierta de K, se debe tener que

$$K \subset \bigcup_{k\in\mathbb{N}} B_{r_k}(\hat{x}_k)$$

y por tanto, que

$$\left(\mathbb{R}^n \setminus \bigcup_{k\in\mathbb{N}} B_{r_k}(\hat{x}_k) \right) \cap K = \varnothing.$$

Con esta contradicción concluimos nuestra prueba. ■

Como seguramente el lector recordará, los conjuntos cerrados y acotados son "viejos conocidos" para nosotros, pues junto con los conjuntos conexos, son el tipo de conjuntos que "preservan" sus características bajo funciones continuas, resultado que dejamos formulado en el teorema 2.4. De esta forma, y como una consecuencia inmediata del teorema de Heine-Borel, podemos reescribir el teorema 2.4 diciendo ahora que los conjuntos compactos también se "preservan" bajo funciones continuas.

Aunque no haría falta hacer ninguna prueba de este hecho, haremos una muy sencilla, en la que sólo usaremos la ("singular") caracterización

de la continuidad de una función dada en la proposición 2.15, y algunas propiedades elementales de la imagen inversa establecidas en la proposición 2.1.

Teorema 2.7 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ continua en A y $K \subset A$. Si K es un conjunto compacto, entonces $f(K)$ es es un conjunto compacto.*

Demostración. Sea $\mathcal{V} = \{V_\alpha \mid \alpha \in I\}$ una cubierta (de abiertos en $\mathbb{R}^m$) de $f(K)$. Por la proposición 2.15, sabemos que para cada $V_\alpha \in \mathcal{V}$ existe $U_\alpha \subset \mathbb{R}^n$ abierto tal que $f^{-1}(V_\alpha) = U_\alpha \cap A$. Ahora, dado que

$$\begin{aligned} K &\subset f^{-1}(f(K)) \\ &\subset f^{-1}\left(\bigcup_{\alpha\in I} V_\alpha\right) \\ &= \bigcup_{\alpha\in I} f^{-1}(V_\alpha) \\ &= \bigcup_{\alpha\in I} (U_\alpha \cap A) \\ &= \left(\bigcup_{\alpha\in I} U_\alpha\right) \cap A \\ &\subset \bigcup_{\alpha\in I} U_\alpha \end{aligned}$$

se tiene que $\mathcal{U} = \{U_\alpha \mid \alpha \in I\}$ es una cubierta abierta de K. Como K es compacto, entonces existen $\alpha_1, \ldots, \alpha_k \in I$ tales que

$$K \subset U_{\alpha_1} \cup \cdots \cup U_{\alpha_k}.$$

Ahora, como $K \subset A$, tenemos que

$$\begin{aligned} K &\subset (U_{\alpha_1} \cup \cdots \cup U_{\alpha_k}) \cap A \\ &= (U_{\alpha_1} \cap A) \cup \cdots \cup (U_{\alpha_k} \cap A) \\ &= f^{-1}(V_{\alpha_1}) \cup \cdots \cup f^{-1}(V_{\alpha_k}) \end{aligned}$$

y por tanto

$$\begin{aligned} f(K) &\subset f\left(f^{-1}(V_{\alpha_1}) \cup \cdots \cup f^{-1}(V_{\alpha_k})\right) \\ &= f(f^{-1}(V_{\alpha_1})) \cup \cdots \cup f(f^{-1}(V_{\alpha_k})) \\ &\subset V_{\alpha_1} \cup \cdots \cup V_{\alpha_k}, \end{aligned}$$

con lo que concluimos que $f(K)$ es compacto. ■

2.4. Problemas

1. Considere las siguientes funciones definidas para algún subconjunto $A \subset \mathbb{R}^2$. En cada caso, identifique al conjunto A y haga un esbozo de su gráfica.

   i) $f(x,y) = x^2 + y^2$ *ii)* $f(x,y) = \sqrt{1-x^2}$

   iii) $f(x,y) = \sqrt{x^2+y^2}$ *iv)* $f(x,y) = xy$

   v) $f(x,y) = \frac{1}{x^2+y^2}$ *vi)* $f(x,y) = x^2 - y^2$

2. Sean $g : A \subset \mathbb{R}^2 \to \mathbb{R}$, $k \in \mathbb{R}$ y $(x_0, y_0) \in \mathbb{R}^2$. Determine el dominio (y describa o esboce la gráfica) de las siguientes funciones en términos del dominio (de la gráfica) de g:

   i) $f(x,y) = g(x,y) + k$ *ii)* $f(x,y) = kg(x,y)$

   iii) $f(x,y) = g(-x,-y)$ *iv)* $f(x,y) = g(x - x_0, y - y_0)$

3. Determine los conjuntos de nivel de cada una de las siguientes funciones:

   i) $f(x,y) = x^2 - y^2$ *ii)* $f(x,y) = 2xy$

   iii) $f(x,y,z) = x^2 - y^2 + z^2$ *iv)* $f(x,y,z) = yz$

4. Determine cuál es el conjunto $f(A)$, si:

a) la función $f : \mathbb{R}^2 \to \mathbb{R}^3$ está definida como

$$f(x, y) = (x\cos(y), x\,\mathrm{sen}(y), x) \quad \text{y} \quad A = \mathbb{R}^2$$

b) la función $f : \mathbb{R}^2 \to \mathbb{R}^3$ está definida como

$$f(x, y) = \left(x^2\cos(y), x^2\,\mathrm{sen}(y), x^2\right) \quad \text{y} \quad A = \mathbb{R}^2$$

c) la función $f : \mathbb{R}^2 \to \mathbb{R}^3$ está definida como

$$f(x, y) = \left(x\cos(y), x\,\mathrm{sen}(y), x^2\right) \quad \text{y} \quad A = \mathbb{R}^2$$

d) la función $f : \mathbb{R}^2 \to \mathbb{R}^3$ está definida como

$$f(x, y) = (a\cos(x), a\,\mathrm{sen}(x), y)\,, a > 0, \quad \text{y} \quad A = [0, 2\pi] \times \mathbb{R} \subset \mathbb{R}^2$$

e) la función $f : \mathbb{R}^2 \to \mathbb{R}^3$ está definida como

$$f(x, y) = (r\cos(x)\,\mathrm{sen}(y), r\,\mathrm{sen}(x)\,\mathrm{sen}(y), r\cos(y))$$

$r > 0$, y $A = [0, 2\pi] \times [-\pi/2, \pi/2] \subset \mathbb{R}^2$

f) la función $f : \mathbb{R}^2 \to \mathbb{R}^3$ está definida como

$$f(x, y) = \left(\sqrt{x^2+1}\cosh(y), \sqrt{x^2+1}\,\mathrm{senh}(y), x\right) \quad \text{y} \quad A = \mathbb{R}^2$$

g) la función $f : \mathbb{R}^2 \to \mathbb{R}^3$ está definida como

$$f(x, y) = \left(x + y, x - y, x^2 - y^2\right) \quad \text{y} \quad A = \mathbb{R}^2.$$

5. Encuentre una función de $\mathbb{R}^2$ en $\mathbb{R}^3$ cuya imagen coincida con el elipsoide

$$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1.$$

6. Considere las siguientes funciones:

a) $f(x, y) = (x\,\mathrm{sen}(y), x\cos(y))$; ¿cuál es la imagen bajo esta función de las rectas de la forma $x = c$ y $y = d$, con c y d cualesquiera números reales?

b) $f(x,y) = (e^x \operatorname{sen}(y), e^x \cos(y))$; ¿cuál es la imagen bajo esta función de las rectas de la forma $x = c$ y $y = d$, con c y d cualesquiera números reales?

c) $f(x,y,z) = (x\cos(y), x\operatorname{sen}(y), z)$; ¿cuál es la imagen bajo esta función de los planos de la forma $x = c$, $y = d$ y $z = k$, con c, d y k cualesquiera números reales?

d) $f(x,y,z) = (x\operatorname{sen}(y)\cos(z), x\operatorname{sen}(y)\operatorname{sen}(z), x\cos(y))$; ¿cuál es la imagen bajo esta función de los planos de la forma $x = c$, $y = d$ y $z = k$, con c, d y k cualesquiera números reales?

7. Sean $g = (g_1, g_2), h : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}^2$ definidas como sigue:

$$g_1(x,y) = \sqrt{x^2+y^2}$$

$$g_2(x,y) = \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{si } x > 0 \\ \frac{\pi}{2} & \text{si } x = 0 \text{ y } y > 0 \\ -\frac{\pi}{2} & \text{si } x = 0 \text{ y } y < 0 \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{si } x < 0 \text{ y } y \geq 0 \\ \arctan\left(\frac{y}{x}\right) - \pi & \text{si } x < 0 \text{ y } y < 0 \end{cases}$$

y $h(x,y) = g(x, y+\pi/2)$, en donde arctan toma sus valores entre $-\pi/2$ y $\pi/2$.

a) determine cuáles son los conjuntos $g(\mathbb{R}^2\setminus\{(0,0)\})$ y $h(\mathbb{R}^2\setminus\{(0,0)\})$

b) pruebe que $(f\circ g)(x,y) = (x,y) = (f\circ h)(x,y)$ para toda $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$

c) si f es la función definida en el inciso (a) del problema 6, ¿para cuáles $(x,y) \in \mathbb{R}^2$ se satisface que $(g\circ f)(x,y) = (x,y)$? ¿para cuáles $(x,y) \in \mathbb{R}^2$ se satisface que $(h\circ f)(x,y) = (x,y)$?

8. Pruebe la proposición 2.1 y de ejemplos de funciones en los que las contenciones de los incisos 7, 8, 9 y 10 sean propias.

9. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}$. Pruebe que para toda $c \in \mathbb{R}$ existe $h : A \subset \mathbb{R}^n \to \mathbb{R}$ tal que $N_c(f) = N_0(h)$.

10. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$.

    a) pruebe que existe $H : A \subset \mathbb{R}^n \to \mathbb{R}^{m+n}$ tal que $G_f = H(A)$.

    b) pruebe que existe $h : U \subset \mathbb{R}^{m+n} \to \mathbb{R}$ tal que $G_f = N_0(h) = h^{-1}(\{0\})$.

11. Decimos que $L : \mathbb{R}^n \to \mathbb{R}^m$ es una función lineal, si $L(\alpha\hat{x} + \beta\hat{y}) = \alpha L(\hat{x}) + \beta L(\hat{y})$ para todos $\alpha, \beta \in \mathbb{R}$, y para todos $\hat{x}, \hat{y} \in \mathbb{R}^n$. Pruebe que:

    a) $L(\hat{0}) = \hat{0}$

    b) si $\hat{c} \in \mathbb{R}^m$, la función $\hat{c} \cdot L$ (ver inciso 3 de la definición 2.1) es una función lineal de $\mathbb{R}^n$ en $\mathbb{R}$

    c) la función constante cero (de $\mathbb{R}^n$ en $\mathbb{R}^m$) es la única función constante que es lineal

    d) existe $M \geq 0$ tal que $\|L(\hat{x})\| \leq M \|\hat{x}\|$ para toda $\hat{x} \in \mathbb{R}^n$

    e) si $L : \mathbb{R}^2 \to \mathbb{R}$ es una función lineal, entonces la gráfica de L es un plano que pasa por el origen

    f) cualquier plano cuya ecuación sea de la forma

    $$Ax + By + Cz = 0$$

    con $C \neq 0$, coincide con ser la gráfica de alguna función lineal $L : \mathbb{R}^2 \to \mathbb{R}$ (de las variables x y y).

    g) si $L : \mathbb{R}^3 \to \mathbb{R}$, con L diferente de la función constante cero, entonces cualquier conjunto de nivel de L es un plano.

12. Determine cuáles de las siguientes sucesiones convergen y cuál es su punto de convergencia (si es que convergen)

i) $\{\hat{x}_k = (k\,\mathrm{sen}(1/k), (1+1/k)^k)\}$

ii) $\{\hat{x}_k = (\mathrm{sen}(k)/k, (1+k)^{1/k})\}$

iii) $\{\hat{x}_k = (a^k + b^k)^{1/k}, kc^k, (-1)^k(1+1/k))\}$

iv) $\{\hat{x}_k = ((k^2+1)^{1/8} - (k+1)^{1/4}, \frac{k}{k+1} - \frac{k+1}{k})\}$

13. Pruebe que una sucesión $\{\hat{x}_k\}$ en $\mathbb{R}^n$ converge al punto $\hat{x}_0 \in \mathbb{R}^n$ si y sólo si la sucesión de números reales $\{\|\hat{x}_k - \hat{x}_0\|\}$ converge a 0.

14. Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$ que converge al punto $\hat{x}_0 \in \mathbb{R}^n$. Pruebe que la sucesión de números reales $\{\|\hat{x}_k\|\}$ converge a $\|\hat{x}_0\|$. ¿Es cierto lo recíproco para cualquier $\hat{x}_0 \in \mathbb{R}^n$?

15. En la definición 2.7 sustituya la norma euclideana por la norma uno y la norma infinito y pruebe que la proposición 2.2 sigue siendo cierta con cada una de ellas. Con base en lo anterior, pruebe que todas estas definiciones de convergencia son equivalentes.

16. Pruebe la proposición 2.3, usando la proposición 2.2 y sin usar dicha proposición.

17. Pruebe la proposición 2.4.

18. En la definición 2.8 sustituya la norma euclideana por la norma uno y la norma infinito y pruebe que la proposición 2.4 sigue siendo cierta con cada una de ellas. Con base en lo anterior, pruebe que todas estas definiciones son equivalentes.

19. Pruebe, sin usar la proposición 2.2, la proposición 2.5 (*sugerencia*: recuerde (o pruebe) que si $\{k_l\}$ es una sucesión creciente de números naturales (es decir que $k_l < k_{l+1}$ para toda $l \in \mathbb{N}$), entonces $l \leq k_l$ para toda $l \in \mathbb{N}$).

20. Sea $\left\{\hat{x}_k = \left(x_k^{(1)}, \ldots, x_k^{(n)}\right)\right\}$ una sucesión en $\mathbb{R}^n$. Pruebe que $\{\hat{x}_k\}$ está acotada si y sólo si $\{x_k^{(i)}\}$ está acotada para cada $i \in \{1, \ldots, n\}$.

21. Pruebe que, si $\{\hat{x}_k\}$ es una sucesión de Cauchy en $\mathbb{R}^n$, entonces cualquier subsucesión $\{\hat{x}_{k_l}\}$ de $\{\hat{x}_k\}$, también es de Cauchy.

22. Sea $\{\hat{x}_k\}$ una sucesión de Cauchy en $\mathbb{R}^n$. Pruebe, directamente de la definición, que:

    a) la sucesión $\{\hat{x}_k\}$ está acotada

    b) si el rango de $\{\hat{x}_k\}$ es finito, entonces existe $N \in \mathbb{N}$ tal que $\hat{x}_k = \hat{x}_0$ para toda $k \geq N$

    c) si $\{\hat{x}_{k_l}\}$ es una subsucesión de $\{\hat{x}_k\}$ que converge a $\hat{x}_0 \in \mathbb{R}^n$, entonces $\{\hat{x}_k\}$ converge a $\hat{x}_0$.

23. Pruebe el corolario 2.1 sin usar la proposición 2.4 (*sugerencia*: para probar que la condición de Cauchy es suficiente para la convergencia, use el problema anterior y el teorema de Bolzano-Weierstrass).

24. Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$ cuyo rango A es finito. Pruebe que existe $\hat{y} \in A$ y una subsucesión $\{\hat{x}_{k_l}\}$ de $\{\hat{x}_k\}$ tal que $\hat{x}_{k_l} = \hat{y}$ para toda $l \in \mathbb{N}$.

25. Sea $\{\hat{x}_k\}$ una sucesión en $\mathbb{R}^n$ cuyo rango A es infinito. Pruebe que:

    a) si $\hat{x}_0$ es un punto de acumulación de A, entonces existe una subsucesión $\{\hat{x}_{k_l}\}$ de $\{\hat{x}_k\}$ que converge a $\hat{x}_0$.

    b) si $\{\hat{x}_k\}$ converge al punto $\hat{x}_0 \in \mathbb{R}^n$, entonces $\hat{x}_0$ es el único punto de acumulación de A.

26. Determine si la siguientes proposiciones son falsas o verdaderas. Pruebe sus respuestas.

    a) si una sucesión $\{\hat{x}_k\}$ en $\mathbb{R}^n$ es tal que su rango tiene un único punto de acumulación, entonces $\{\hat{x}_k\}$ es convergente

    b) existe una sucesión $\{\hat{x}_k\}$ en $\mathbb{R}^n$ no convergente que tiene una única subsucesión convergente.

27. Pruebe, usando el teorema de Bolzano-Weierstrass y sin usar el caso real, el teorema 2.1 (*sugerencia*: analice dos casos: cuando el rango de $\{\hat{x}_k\}$ es finito, en cuyo caso será útil el problema 24, y cuando el rango de $\{\hat{x}_k\}$ es infinito, en cuyo caso será útil el teorema de Bolzano-Weierstrass y el inciso (b) del problema 25).

28. Pruebe, sin usar el teorema de Bolzano-Weierstrass y el caso real, el teorema 2.1 (*sugerencia*: use que los naturales son infinitos y el teorema de los rectángulos anidados, "imitando" el método de la cacería del león usado en la prueba del teorema de Bolzano-Weierstrass).

29. Sea $A \subset \mathbb{R}^n$. Pruebe que:

    a) $\hat{x}_0 \in \bar{A}$ si y sólo si existe una sucesión $\{\hat{x}_k\}$ en A tal que $\{\hat{x}_k\}$ converge a $\hat{x}_0$

    b) $\hat{x}_0 \in A'$ si y sólo si existe una sucesión $\{\hat{x}_k\}$ en $A \setminus \{\hat{x}_0\}$ tal que $\{\hat{x}_k\}$ converge a $\hat{x}_0$

    c) si $\hat{x}_0$ es un punto aislado de A y $\{\hat{x}_k\} \subset A$ es una sucesión que converge a $\hat{x}_0$, entonces existe $N \in \mathbb{N}$ tal que $\hat{x}_k = \hat{x}_0$ para toda $k \geq N$.

30. Pruebe la proposición 2.7.

31. Sea $f : A \subset \mathbb{R} \to \mathbb{R}^n$ y $x_0 \in \mathbb{R}$.

    a) Defina el concepto de límite lateral (por la izquierda y por la derecha) de f en x_0; es decir, defina lo que significa que $\hat{l} \in \mathbb{R}^n$ sea tal que

    $$\hat{l} = \lim_{x \to x_0^+} f(x)$$

    o que

    $$\hat{l} = \lim_{x \to x_0^-} f(x)$$

    Para dar estas definiciones, ¿qué condición es necesario que cumpla el punto x_0? ¿es suficiente que $x_0 \in A'$? Justifique su respuesta.

b) Con base en las definiciones anteriores (y suponiendo que x_0 satisface las condiciones que hayan hecho falta), pruebe que

$$\hat{l} = \lim_{x \to x_0} f(x) \quad \text{si y sólo si} \quad \hat{l} = \lim_{x \to x_0^+} f(x) \quad \text{y} \quad \hat{l} = \lim_{x \to x_0^-} f(x).$$

32. Determine si las siguientes funciones tienen límite en el punto que se indica. Pruebe su respuesta.

i) $f(x,y) = \frac{x^3y^2}{x^6+y^4}$ en $(0,0)$

ii) $f(x,y) = \frac{x^3+xy^2-x^2y-y^3}{x^2+y^2}$ en $(0,0)$

iii) $f(x,y) = \frac{x^2(y-1)}{x^4+(y-1)^2}$ en $(0,1)$

iv) $f(x,y) = \frac{x^2y+x^2}{x^2+y^2+2y+1}$ en $(0,-1)$

v) $f(x,y,z) = \frac{xy+yz+zx}{x^2+y^2+z^2}$ en $(0,0,0)$

vi) $f(x,y,z) = \frac{xy+yz+zx}{\sqrt{x^2+y^2+z^2}}$ en $(0,0,0)$

vii) $f(x,y) = \frac{x^3y^4}{x^4+y^4}$ en $(0,0)$

viii) $f(x,y) = \frac{xy^3}{x^2+y^6}$ en $(0,0)$

ix) $f(x,y) = \frac{x^3+y^4}{x^2+y^2}$ en $(0,0)$

x) $f(x,y) = \frac{x^2y^5}{x^2+(y-1)^2}$ en $(0,1)$

33. Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $\hat{x}_0 \in A'$. Pruebe que $\lim_{\hat{x} \to \hat{x}_0} f(\hat{x}) = \hat{0}$ si y sólo si $\lim_{\hat{x} \to \hat{x}_0} \|f(\hat{x})\| = 0$.

34. Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $\hat{x}_0 \in \mathbb{R}^n$ y $\hat{l} \in \mathbb{R}^m$. Definimos $B =$

$\{\hat{x} - \hat{x}_0 \in \mathbb{R}^n \mid \hat{x} \in A\}$ y $g : B \subset \mathbb{R}^n \to \mathbb{R}^m$ como $g(\hat{h}) = f(\hat{x}_0 + \hat{h})$ para cada $\hat{h} \in B$. Pruebe que:

a) $\hat{x}_0 \in A'$ si y sólo si $\hat{0} \in B'$

b) $\hat{l} = \lim_{\hat{x}\to\hat{x}_0} f(\hat{x})$ si y sólo si $\hat{l} = \lim_{\hat{h}\to\hat{0}} g(\hat{h})$.

35. Sean $f, g : A \subset \mathbb{R}^n \to \mathbb{R}^m$ y $\hat{x}_0 \in A'$. Pruebe que, si g está acotada en una vecindad "agujerada" de $\hat{x}_0$ (es decir que existen $r > 0$ y $M > 0$ tales que $\|g(\hat{x})\| \leq M$ para toda $\hat{x} \in (B_r(\hat{x}_0) \setminus \{\hat{x}_0\}) \cap A$) y $\lim_{\hat{x}\to\hat{x}_0} f(\hat{x}) = \hat{0}$, entonces $\lim_{\hat{x}\to\hat{x}_0}(g \cdot f)(\hat{x}) = 0$.

36. En la definición 2.13 sustituya la norma euclideana por las normas uno e infinito y pruebe que todas estas definiciones son equivalentes.

37. Sean $f, g : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $\hat{x}_0 \in A'$ y $k \in \mathbb{N}$. Pruebe que, si

$$\lim_{\hat{x}\to\hat{x}_0} \frac{\|f(\hat{x}) - g(\hat{x})\|}{\|\hat{x} - \hat{x}_0\|^k} = 0,$$

entonces

$$\lim_{\hat{x}\to\hat{x}_0} \frac{\|f(\hat{x}) - g(\hat{x})\|}{\|\hat{x} - \hat{x}_0\|^s} = 0$$

para toda $s \in \mathbb{Z}$ tal que $s \leq k$.

38. Sea $L : \mathbb{R}^n \to \mathbb{R}^m$ una función lineal, es decir que $L(\alpha\hat{x} + \beta\hat{y}) = \alpha L(\hat{x}) + \beta L(\hat{y})$ para todos $\alpha, \beta \in \mathbb{R}$, y para todos $\hat{x}, \hat{y} \in \mathbb{R}^n$. Pruebe que, si

$$\lim_{\hat{h}\to\hat{0}} \frac{\left\|L(\hat{h})\right\|}{\left\|\hat{h}\right\|} = 0,$$

entonces L es la función constante cero ($L \equiv 0$).

39. Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, $\hat{x}_0 \in A'$ y $\hat{l} \in \mathbb{R}^m$ tales que $\hat{l} = \lim_{\hat{x}\to\hat{x}_0} f(\hat{x})$. Determine si las siguientes afirmaciones son ciertas. Pruebe sus respuestas.

a) $\hat{l} \in (f(A))'$

b) si $D \subset \mathbb{R}^m$ es tal que $\hat{l} \in D'$ y $(f^{-1}(D))' \neq \varnothing$, entonces $\hat{x}_0 \in (f^{-1}(D))'$

c) si las afirmaciones anteriores no son ciertas, dé hipótesis adicionales sobre la función f para que éstas sí sean ciertas

40. Pruebe la proposición 2.9, primero usando la proposición 2.3, y después usando sólo la definición 2.13.

41. Pruebe las proposiciones 2.10, 2.11 y 2.12.

42. Determine en qué puntos de su dominio son continuas las funciones g y h definidas en el problema 7. Pruebe su respuesta.

43. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ continua en $\hat{x}_0 \in A$ tal que $f(\hat{x}_0) \neq \hat{0}$. Pruebe que:

a) existe $\delta > 0$ tal que $f(\hat{x}) \neq \hat{0}$ para toda $\hat{x} \in B_\delta(\hat{x}_0) \cap A$

b) existen $c > 0$ y $\delta' > 0$ tales que $\|f(\hat{x})\| \geq c$ para toda $\hat{x} \in B_{\delta'}(\hat{x}_0) \cap A$.

44. Pruebe la proposición 2.14.

45. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$. Pruebe que f es continua en A si y sólo si para todo conjunto cerrado $C \subset \mathbb{R}^m$ existe un conjunto cerrado $D \subset \mathbb{R}^n$ tal que $f^{-1}(C) = D \cap A$.

46. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$. Pruebe que:

a) f es continua en A si y sólo si $f\left(\bar{B} \cap A\right) \subset \overline{f(B)}$ para todo $B \subset A$

b) f es continua en A si y sólo si $f\left(B' \cap A\right) \subset \overline{f(B)}$ para todo $B \subset A$.

47. Pruebe que:

a) $A = \{(x, y) \in \mathbb{R}^2 \mid 1 < x^2 + y\}$ es un conjunto abierto

b) $A = \{(x, y, z) \in \mathbb{R}^3 \mid (zx + zy)/(x^2 + y^2) < 0\}$ es un conjunto abierto

c) $A = \{(x, y) \in \mathbb{R}^2 \mid y = 1/x\}$ es un conjunto cerrado.

48. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $(a, b) \subset \mathbb{R}$ tal que $f(U) \subset (a, b)$ y $g : (a, b) \subset \mathbb{R} \to \mathbb{R}$ derivable en $f(\hat{x}_0)$. Definimos $\varphi : U \subset \mathbb{R}^n \to \mathbb{R}$ como

$$\varphi(\hat{x}) = \begin{cases} \frac{g(f(\hat{x})) - g(f(\hat{x}_0))}{f(\hat{x}) - f(\hat{x}_0)} & \text{si } f(\hat{x}) - f(\hat{x}_0) \neq 0 \\ \\ g'(f(\hat{x}_0)) & \text{si } f(\hat{x}) - f(\hat{x}_0) = 0 \end{cases}$$

Pruebe que, si f es continua en $\hat{x}_0$, entonces φ es continua en $\hat{x}_0$.

49. Pruebe que las siguientes funciones son continuas en su dominio:

a) $f : \mathbb{R}^n \to \mathbb{R}$ definida como $f(\hat{x}) = \|\hat{x}\|$

b) $f : \mathbb{R}^n \to \mathbb{R}$ definida como $f(\hat{x}) = f(x_1, \ldots, x_n) = x_i$, donde $i \in \{1, \ldots, n\}$

c) $L : \mathbb{R}^n \to \mathbb{R}^m$ cualquier función lineal

50. Sea

$$S^{n-1} = \{\hat{x} \in \mathbb{R}^n \mid \|\hat{x}\| = 1\}.$$

Pruebe que S^{n-1} es un conjunto cerrado.

51. Sean, $A \subset \mathbb{R}^n$ un abierto, $\hat{x} \in A, \hat{y} \in (A \cup A')^c$ y $f : [0, 1] \subset \mathbb{R} \to \mathbb{R}^n$ continua tal que $f(0) = \hat{x}$ y $f(1) = \hat{y}$. Pruebe que existe $t \in (0, 1)$ tal que $f(t) \in Fr(A)$.

52. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}$ continua en A, con A conexo y tal que $f(\hat{x}) \neq 0$ para toda $\hat{x} \in A$. Pruebe que $f(\hat{x}) > 0$ para toda $\hat{x} \in A$ o $f(\hat{x}) < 0$ para toda $\hat{x} \in A$.

53. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ continua en A, con A conexo y tal que $\|f(\hat{x})\| \neq 1$ para toda $\hat{x} \in A$. Pruebe que si $\|f(\hat{x}_0)\| < 1$ para alguna $\hat{x}_0 \in A$, entonces $\|f(\hat{x})\| < 1$ para toda $\hat{x} \in A$.

54. Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}$ continua en A, y $B \subset A$ conexo, cerrado y acotado. Pruebe que existen $a, b \in \mathbb{R}$ tales que $f(B) = [a, b]$.

55. Sea $A \subset \mathbb{R}^n$ un conjunto no vacío, cerrado y acotado, y $\hat{y} \in A^c$. Pruebe que existe $\hat{x}_0 \in A$ tal que $\|\hat{y} - \hat{x}_0\| \leq \|\hat{x} - \hat{y}\|$ para todo $\hat{x} \in A$. Muestre, con un ejemplo, que esta afirmación no es válida si no suponemos que A es cerrado. ¿Esta afirmación sigue siendo válida si sólo suponemos que A es cerrado? Pruebe su respuesta.

56. Sean $A, B \subset \mathbb{R}^n$ conjuntos no vacíos, cerrados y acotados, tales que $A \cap B = \varnothing$. Pruebe que existen $\hat{x}_0 \in A$ y $\hat{y}_0 \in B$ tales que $\|\hat{y}_0 - \hat{x}_0\| \leq \|\hat{x} - \hat{y}\|$ para todo $\hat{x} \in A$ y para todo $\hat{y} \in B$. Muestre, con un ejemplo, que esta afirmación no es válida si no suponemos que A es cerrado.

57. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ continua, con A cerrado y acotado. Pruebe que existen $\hat{x}_0, \hat{x}_1 \in A$ tales que $\|f(\hat{x}_0)\| \leq \|f(\hat{x})\| \leq \|f(\hat{x}_1)\|$ para toda $\hat{x} \in A$.

58. Sea $K \subset \mathbb{R}^n$. Pruebe que el conjunto K es compacto si y sólo si toda sucesión $\{\hat{x}_k\} \subset K$ tiene una subsucesión $\{\hat{x}_{k_l}\}$ que converge a un punto $\hat{x}_0 \in K$.

59. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ continua e inyectiva en A, con A cerrado y acotado. Pruebe que $f^{-1} : f(A) \subset \mathbb{R}^m \to \mathbb{R}^n$ (la función inversa de f) es continua en $f(A)$. ¿Esta afirmación se sigue cumpliendo si A no es cerrado? Pruebe su respuesta.

60. En la definición 2.16 sustituya la norma euclideana por las normas uno e infinito y pruebe que todas estas definiciones son equivalentes.

61. Pruebe la proposición 2.16.

62. Pruebe que la función definida en el ejemplo 2.13 es uniformemente continua sobre cualquier subconjunto $B \subset \mathbb{R}^2 \setminus \{(0, 0)\}$ que no tenga como punto de acumulación al $(0, 0)$.

63. Sean $A \subset \mathbb{R}^n$ cerrado y acotado, y $f : [a,b] \times A \subset \mathbb{R}^{n+1} \to \mathbb{R}$ continua en $[a,b] \times A \subset \mathbb{R}^{n+1}$. Definimos $h : A \subset \mathbb{R}^n \to \mathbb{R}$ como

$$h(\hat{y}) = \int_a^b f(x, \hat{y}) dx$$

Pruebe que h es uniformemente continua en A.

64. Sea $L : \mathbb{R}^n \to \mathbb{R}^m$ una función lineal. Pruebe que L es uniformemente continua en $\mathbb{R}^n$.

65. Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ uniformemente continua en A y $\{\hat{x}_k\} \subset A$ una sucesión de Cauchy. Pruebe que $\{f(\hat{x}_k)\}$ es una sucesión de Cauchy.

66. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ uniformemente continua en A. Pruebe que existe $\tilde{f} : \bar{A} \subset \mathbb{R}^n \to \mathbb{R}^m$ uniformemente continua en $\bar{A}$ tal que $\tilde{f}(\hat{x}) = f(\hat{x})$ para toda $\hat{x} \in A$.

67. Sea $A \subset \mathbb{R}^n$ no vacío. Definimos $f_A : \mathbb{R}^n \to \mathbb{R}$ como

$$f_A(\hat{x}) = dist(\hat{x}, A) := \text{ínf}\{\|\hat{x} - \hat{y}\| \mid \hat{y} \in A\}$$

(para cada $\hat{x} \in \mathbb{R}^n$). Pruebe que:

   a) $f_A(\hat{x}) = 0$ si y sólo si $\hat{x} \in \bar{A}$

   b) f_A es uniformemente continua en $\mathbb{R}^n$ (*sugerencia*: pruebe que $|f_A(\hat{x}) - f_A(\hat{y})| \leq \|\hat{x} - \hat{y}\|$ para todo $\hat{x}, \hat{y} \in \mathbb{R}^n$)

   c) si $A, B \subset \mathbb{R}^n$ están separados, entonces existen $U, V \subset \mathbb{R}^n$ abiertos tales que $A \subset U$, $B \subset V$ y $U \cap V = \emptyset$ (*sugerencia*: considere la función $f(\hat{x}) = f_A(\hat{x}) - f_B(\hat{x})$)

   d) ¿La afirmación del problema 56 sigue siendo válida si sólo suponemos que A es cerrado? Pruebe su respuesta.

Capítulo 3

La derivada de funciones de $\mathbb{R}$ en $\mathbb{R}^n$

Con este breve capítulo damos inicio al estudio del concepto de derivación para funciones de varias variables, pero a diferencia del capítulo anterior, empezaremos haciéndolo sólo para el caso de funciones de $\mathbb{R}$ en $\mathbb{R}^n$.

Como también mencionamos anteriormente, dado que lo importante de las funciones de este tipo es su imagen, éstas suelen ser útiles para describir "objetos geométricos" a los que (bajo ciertas condiciones) nos referiremos como "*curvas*"; o para describir el "movimiento" de un objeto, razón por la cual a estas funciones también las conoceremos con el nombre de "*trayectorias*". Justo a partir de estos dos "usos" es que "motivaremos" su concepto de derivada, pero previamente daremos algunos ejemplos de ambas formas de usarlas.

3.1. Geometría y movimiento

Aun cuando a estas alturas el lector posiblemente no conozca una definición precisa de lo que significa la palabra "*curva*" (definición que precisaremos más adelante), cualquiera que ésta fuera debiera de abarcar a "objetos geométricos" tan conocidos como las rectas y las *cónicas* en el plano. A fal-

ta de tal definición y a manera de ejemplo, por ahora sólo nos limitaremos a mostrar que, entre otros muchos "objetos geométricos", las rectas y las cónicas se pueden obtener como la imagen de una función de $\mathbb{R}$ en $\mathbb{R}^2$. Tal es el caso de las rectas en el plano ($\mathbb{R}^2$), las cuales podemos pensar en general como un conjunto definido de la siguiente forma:

$$R = \{(x, y) \in \mathbb{R}^2 \mid ax + by + c = 0\},$$

en donde $a^2 + b^2 > 0$.

Una forma muy sencilla de ver a R como la imagen de una función de $\mathbb{R}$ en $\mathbb{R}^2$, consiste en observar que, como $a^2 + b^2 > 0$, entonces $a \neq 0$ o $b \neq 0$, de modo que si suponemos que sucede lo primero, entonces para cualquier $(x, y) \in R$ se tiene que

$$x = \frac{-c - by}{a}. \tag{3.1}$$

Por tanto, a la pareja (x, y) la podemos escribir como

$$\left(\frac{-c - by}{a}, y\right),$$

es decir, la podemos escribir en términos de una sola variable (o de un sólo *parámetro*, que es el témino que se suele usar en este contexto). Basados en lo anterior, es fácil comprobar que si consideramos la función[1] $\gamma : \mathbb{R} \to \mathbb{R}^2$ dada por

$$\gamma(t) = \left(\frac{-c - bt}{a}, t\right),$$

se tiene que $\gamma(\mathbb{R}) = R$.

En efecto, nótese que para toda $t \in \mathbb{R}$ se tiene que

$$\begin{aligned} a\left(\frac{-c - bt}{a}\right) + bt + c &= (-c - bt) + bt + c \\ &= 0, \end{aligned}$$

[1]Es común que se usen letras griegas para nombrar a las funciones de $\mathbb{R}$ en $\mathbb{R}^n$.

con lo cual concluimos que $\gamma(\mathbb{R}) \subset R$. Por otra parte, si $(x, y) \in R$, por la identidad 3.1 tenemos que bastará tomar $t = y$ para que se cumpla que

$$\begin{aligned}\gamma(y) &= \left(\frac{-c - by}{a}, y\right) \\ &= (x, y),\end{aligned}$$

con lo que concluimos que $R \subset \gamma(\mathbb{R})$ y por lo tanto que, $\gamma(\mathbb{R}) = R$.

Aunque laborioso (pero no difícil), también se puede probar que la circunferencia con centro en el origen (de un sistema coordenado cartesiano dado) de radio $r > 0$ dada por

$$C_r = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = r^2\}$$

se puede obtener como la imagen de la función $\gamma : [0, 2\pi) \to \mathbb{R}^2$ dada por

$$\gamma(t) = (r\cos(t), r\,\mathrm{sen}(t)).$$

O que la elipse $E \subset \mathbb{R}^2$ dada por

$$E = \left\{(x, y) \in \mathbb{R}^2 \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\right\}$$

se puede obtener como la imagen de la función $\gamma : [0, 2\pi) \to \mathbb{R}^2$ dada por

$$\gamma(t) = (a\cos(t), b\,\mathrm{sen}(t)).$$

O que la parábola $P \subset \mathbb{R}^2$ dada por

$$P = \{(x, y) \in \mathbb{R}^2 \mid y = x^2\}$$

se puede obtener como la imagen de la función $\gamma : \mathbb{R} \to \mathbb{R}^2$ dada por

$$\gamma(t) = (t, t^2).$$

Aun cuando todavía no tenemos todo lo necesario para definir lo que significa que un subconjunto de $\mathbb{R}^n$ sea *una curva*, con base en estos ejemplos ya podemos definir lo que significa que una de estas funciones *parametrice* a un conjunto $C \subset \mathbb{R}^n$.

En general, si $C \subset \mathbb{R}^n$ es tal que coincide con la imagen de una función de $\mathbb{R}$ en $\mathbb{R}^n$, diremos que dicha función es *una parametrización* de C.

Definición 3.1 *Sea $C \subset \mathbb{R}^n$. Si existe $\gamma = (\gamma_1, \ldots, \gamma_n) : I \subset \mathbb{R} \to \mathbb{R}^n$ tal que $\gamma(I) = C$ decimos que γ es una parametrización de C. En este caso diremos que las identidades*

$$\begin{aligned} x_1 &= \gamma_1(t) \\ x_2 &= \gamma_2(t) \\ &\vdots \\ x_n &= \gamma_n(t) \end{aligned}$$

son unas ecuaciones paramétricas de C.

Con relación a la definición anterior, es importante llamar la atención sobre el hecho de que a la función γ no se le pide ninguna *propiedad*. Identificar los subconjuntos de $\mathbb{R}^n$ para los cuales existe una parametrización es sin duda un problema interesante (y tal vez no muy difícil), pero que está fuera de los objetivos de este texto. De hecho, hablando de propiedades de funciones de $\mathbb{R}$ en $\mathbb{R}^n$, hemos definido ya lo que significa que una de estas funciones sea continua, y sin duda otra pregunta interesante sería la siguiente: ¿qué subconjuntos de $\mathbb{R}^n$ tienen una parametrización que sea continua? Este es un problema mucho mas interesante (¡y difícil!) que el anterior, y a los deseosos en saber algo al respecto les recomendamos investigar acerca de las llamadas *curvas de Peano*[2].

Como seguramente el lector ya habrá notado, más que responder estas preguntas, el objetivo de este texto está centrado en desarrollar las propiedades que pueden tener este tipo de funciones, entre la cuales se encuentra la de la derivabilidad.

Con respecto a la misma definición 3.1, también vale la pena destacar el uso que se hace del artículo "*una*". En efecto, nótese que si $I \subset \mathbb{R}$ y γ son como en dicha definición, y se tiene que $\alpha : J \subset \mathbb{R} \to \mathbb{R}$ es tal que $\alpha(J) = I$, entonces $\tilde{\gamma} : J \subset \mathbb{R} \to \mathbb{R}^n$ definida como $\tilde{\gamma} = \gamma \circ \alpha$ es tal que $\tilde{\gamma}(J) = C$ y por tanto $\tilde{\gamma}$ será "*otra*" parametrización de C. Es decir, si

[2]Giuseppe Peano (Spinetta, 27 de agosto de 1858 – Turín, 20 de abril de 1932) fue un matemático, lógico y filósofo italiano, conocido por sus contribuciones a la lógica matemática y la teoría de números.

un subconjunto $C \subset \mathbb{R}^n$ tiene una parametrización, entonces con base en ella podemos obtener más parametrizaciones del mismo conjunto. Por esta misma razón, lo más adecuado es decir que

$$\begin{aligned} x &= r\cos(t) \\ y &= r\,\mathrm{sen}(t) \end{aligned}$$

para $t \in [0, 2\pi)$, son "*unas*" ecuaciones paramétricas de la circunferencia de radio $r > 0$ con centro en el origen, y que

$$\begin{aligned} x &= r\,\mathrm{sen}(2\pi t) \\ y &= r\cos(2\pi t) \end{aligned}$$

para $t \in [0, 1)$, son "*otras*" ecuaciones paramétricas de la misma circunferencia.

Otro tipo de "objeto geométrico" que se puede obtener como la imagen de una función de $\mathbb{R}$ en $\mathbb{R}^2$ es uno con el cual el lector debe estar muy familiarizado: la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$. En efecto, si $f : I \subset \mathbb{R} \to \mathbb{R}$, sabemos que la gráfica de f es un subconjunto de $\mathbb{R}^2$ definido como

$$G_f := \big\{(x, f(x)) \in \mathbb{R}^2 \mid x \in I\big\}$$

y es muy fácil comprobar que la función $\gamma : I \subset \mathbb{R} \to \mathbb{R}^2$ dada por $\gamma(t) = (t, f(t))$ es una parametrización de G_f.

Seguramente el lector estará de acuerdo en que la observación anterior incrementa sustancialmente la cantidad de subconjuntos de $\mathbb{R}^2$ que podemos describir como la imagen de una función de $\mathbb{R}$ en $\mathbb{R}^2$. Ahora sabemos que subconjuntos como las gráficas de la funciones $f(x) = |x|$ o $f(x) = \mathrm{sen}(x)$ se pueden obtener de esta forma (ver figura 3.1).

Concluimos esta breve sección mostrando cómo las funciones de $\mathbb{R}$ en $\mathbb{R}^n$ resultan ser una herramienta adecuada para describir el "camino" o "trayectoria" seguido por un objeto. Para ello, recurriremos a un conocido ejemplo que justamente se puede plantear en estos términos.

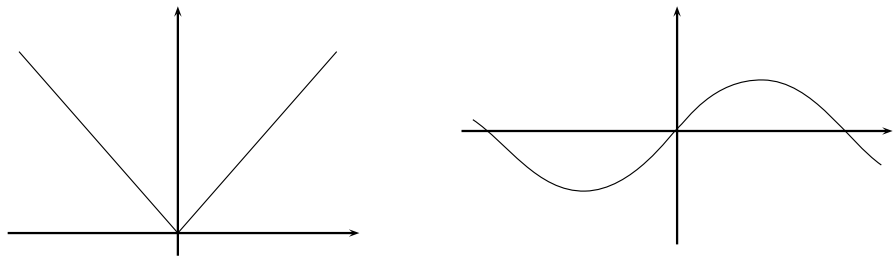


Figura 3.1: Las gráficas de la funciones $f(x) = |x|$ y $f(x) = \operatorname{sen}(x)$ se pueden obtener como la imagen de una función de $\mathbb{R}$ en $\mathbb{R}^2$.

Suponga que se tiene una rueda, la cual vamos a rodar (sin resbalar) en línea recta y sobre una superficie plana, y sobre dicha rueda hacemos una marca. Nuestro objetivo es encontrar una función que nos proporcione la posición (una vez establecido un sistema cartesiano de referencia) de dicha marca, conforme la rueda va girando. Para ello, simplificaremos el problema sustituyendo la rueda por una circunferencia de radio $r > 0$, y a la marca por un punto P de dicha circunferencia. Supondremos que la posición "inicial" de la circunferencia es tal que su centro C se encuentra en el punto de coordenadas $(0, r)$ y que las correspondientes al punto P están dadas por $(0, 2r)$ (ver figura 3.2 (a)). También supondremos que la circunferencia rueda hacia la derecha, en la dirección positiva del eje X.

Una cuestión muy importante en este tipo de problemas es la elección de la variable (o parámetro, que es el término que se suele usar en estos casos) en términos de la cual queremos expresar la posición del punto que nos interesa. Esta elección depende en general del problema que se esté tratando (lo que significa que no existen "reglas" para su elección).

En el caso que nos ocupa, una vez que la circunferencia haya rodado una cierta distancia (que supondremos fue pequeña), el punto P, cuyo movimiento nos interesa describir, ocupará una nueva posición; las coordenadas de esta nueva posición las vamos a determinar usando el ángulo formado por la semirecta (paralela al eje Y) que parte del centro de la circuferencia (en dirección hacia arriba), y el segmento de recta que une al centro C

con el punto P, ángulo que denotaremos por la letra θ y que de acuerdo a la observación que hicimos antes, jugará el papel de parámetro (ver figura 3.2 (b)).

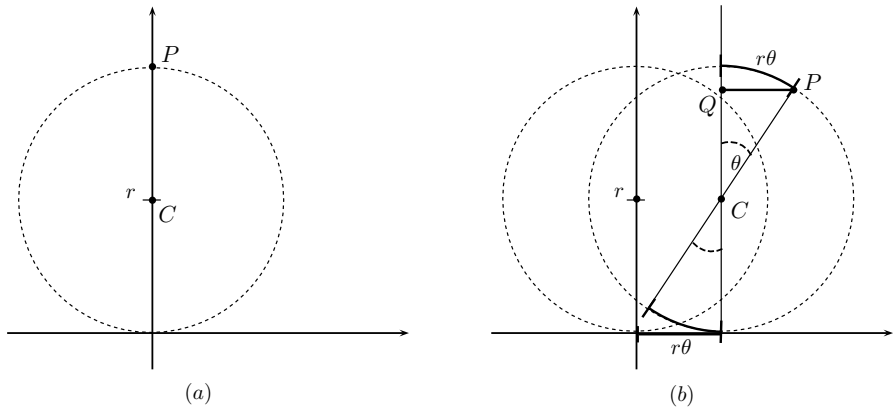


Figura 3.2: Deducción de una parametrización de la cicloide.

Lo primero que haremos será encontrar las nuevas coordenadas del centro C de la circunferencia desplazada. Para ello es importante hacer notar que la distancia recorrida por dicho centro, dado que la circunferencia rueda sin resbalar, coincide con la longitud del arco subtendido por el ángulo opuesto a θ, la cual está dada por $r\theta$. De esta forma, las coordenadas del centro C son $(r\theta, r)$.

Si ahora observamos que para obtener las coordenadas del punto P, basta con tomar las de punto C y sumarle la longitud de los correspondientes catetos del triángulo rectángulo formado por los puntos C, P y Q (ver figura 3.2 (b)), concluimos que las nuevas coordenadas de P están dadas por $(r\theta + r\,\mathrm{sen}(\theta), r + r\cos(\theta))$. Por tanto, la función de $\mathbb{R}$ en $\mathbb{R}^2$ definida como

$$\gamma(\theta) = r(\theta + \mathrm{sen}(\theta), 1 + \cos(\theta))$$

describe el movimiento del punto P en términos del ángulo θ. En la figura 3.3 se muestra la imagen de esta función para $\theta \in [0, 2\pi]$ y representa a la

"trayectoria" seguida por el punto P cuando la circunferencia ha realizado una vuelta completa, "trayectoria" conocida con el nombre de *cicloide*.

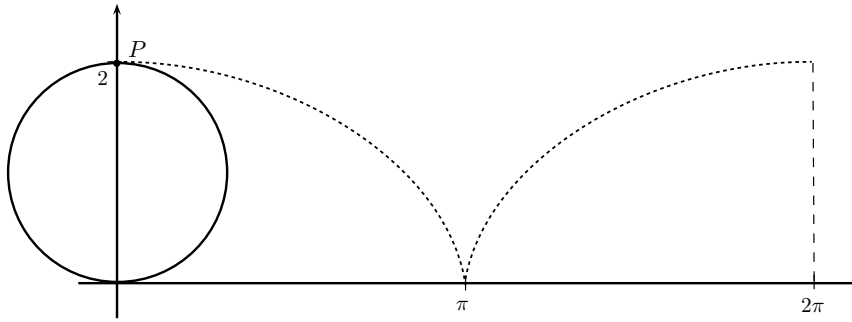


Figura 3.3: Trayectoria seguida por el punto P cuando la circunferencia (de radio 1) ha realizado una vuelta completa, trayectoria conocida con el nombre de cicloide.

Existen una buena cantidad de interesantes "trayectorias" definidas de manera análoga, pero desafortunadamente no está entre los objetivos de este libro profundizar en este tipo de ejemplos[3]. Nuestro objetivo principal en este capítulo es introducir el concepto de derivada para funciones de $\mathbb{R}$ en $\mathbb{R}^n$, y lo motivaremos a partir del uso de éstas en la geométría y en la física, razón por la cual analizamos sólo estos pocos ejemplos.

3.2. La derivada

Dada una función $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ con la cual "describimos" a un conjunto $C \subset \mathbb{R}^n$, si éste tiene el aspecto de una "línea doblada" (como por ejemplo una cónica), nos vamos a plantear el problema de encontrar la "recta tangente" a C en un punto $\hat{x}_0$.

Si $t_0 \in I$ es tal que $\gamma(t_0) = \hat{x}_0$ y $t \in I$ es "cercano" a t_0, ver a la diferencia

$$\gamma(t) - \gamma(t_0) \tag{3.2}$$

[3]Para el lector interesado en conocer más ejemplos de este tipo, le recomendamos visitar el sitio de internet: www-history.mcs.st-and.ac.uk/Curves/Curves.html

como una "flecha" nos da la oportunidad de obtener la recta que pasa por $\hat{x}_0$ y que sigue la dirección determinada por $\gamma(t) - \gamma(t_0)$ (ver problema 3) suponiendo, por supuesto, que esta flecha no es el vector $\hat{0}$.

Como se muestra en la figura 3.4, esta recta es secante a la curva C y en la medida de que t esté más "cercano" a t_0, dicha secante se verá cada vez más como una tangente.

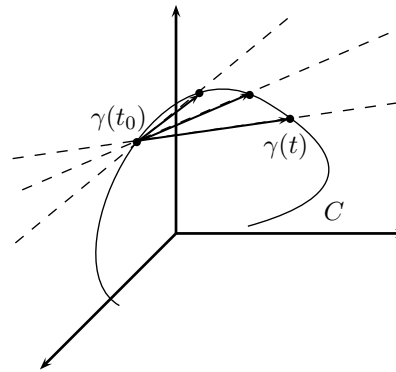


Figura 3.4: Si el vector $\gamma(t) - \gamma(t_0)$ no es el vector $\hat{0}$, la recta cuya dirección está determinada por éste y que pasa por el punto $\gamma(t_0)$ es una recta secante a la curva C que "tiende" a ser tangente (en $\gamma(t_0)$) si t tiende a t_0.

La construcción anterior estaría muy bien si no fuera por el hecho de que la diferencia 3.2 se parecerá cada vez más al vector $\hat{0}$ en la medida de que t esté más "cercano" a t_0 (sobre todo si γ es una función continua, lo que seguramente sucederá en la mayoría de los casos que sean de nuestro interés), con lo cual perdemos toda oportunidad de definir una recta.

La "solución" que daremos a este problema será que, en lugar de considerar al vector 3.2, vamos a considerar al vector dado por la expresión

$$\frac{1}{t - t_0}(\gamma(t) - \gamma(t_0)) = \frac{\gamma(t) - \gamma(t_0)}{t - t_0}, \tag{3.3}$$

con el cual, si no es el vector $\hat{0}$, podemos seguir construyendo rectas secantes a C.

Una razón por la cual vamos a considerar al vector 3.3 es que éste no sólo nos sigue proporcionando información de cuál es la "dirección que tiende" a ser tangente al conjunto C en el punto $\hat{x}_0$, sino que además su norma nos da información sobre la razón de cambio (con respecto a la distancia entre t y t_0) de la distancia entre los puntos $\gamma(t)$ y $\gamma(t_0)$.

Más adelante, cuando veamos a γ como una función que describe el movimiento de un objeto, también veremos que el vector 3.3 es la forma más adecuada de medir la "velocidad promedio" de dicho objeto y, en última instancia, veremos cómo este vector se relaciona de manera muy "adecuada" con el concepto de derivada para funciones de $\mathbb{R}$ en $\mathbb{R}$, justo del tipo que son las funciones coordenadas de γ.

Como seguramente el lector ya intuye, lo siguiente que haremos será fijarnos en el

$$\lim_{t\to t_0} \frac{\gamma(t)-\gamma(t_0)}{t-t_0}.$$

En caso de que éste exista (y que no sea el vector $\hat{0}$, lo que no siempre podemos asegurar), si a dicho límite lo denotamos (casualmente) por $\gamma'(t_0)$, sin duda que la recta definida paramétricamente por la expresión

$$\gamma(t_0)+t\gamma'(t_0)$$

con $t\in\mathbb{R}$, es una recta que tiene el aspecto y las características adecuadas para ser llamada "la recta tangente" a C en el punto $\hat{x}_0=\gamma(t_0)$.

Con base en la discusión anterior, damos las siguientes definiciones.

Definición 3.2 *Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$, con I un intervalo. Dado $t_0 \in I$, decimos que γ es derivable en t_0 si*

$$\lim_{t\to t_0} \frac{\gamma(t)-\gamma(t_0)}{t-t_0}$$

existe. Este límite lo denotamos por $\gamma'(t_0)$ y lo llamamos la derivada de γ en t_0. Es decir,

$$\gamma'(t_0) := \lim_{t\to t_0} \frac{\gamma(t)-\gamma(t_0)}{t-t_0}.$$

Observación 3.1 *En virtud de que restringimos la definición anterior a funciones cuyo dominio es un intervalo, es importante señalar que en caso de que $t_0 \in I$ fuera un "extremo" de I, el límite que se deberá tomar será el correspondiente límite lateral, cuya definición fue dada por el lector en el problema 31 del capítulo 2.*

Definición 3.3 *Sea $C \subset \mathbb{R}^n$ y $\hat{x}_0 \in C$. Si $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ es una parametrización de C tal que $\gamma(t_0) = \hat{x}_0$ para alguna $t_0 \in I$ y $\gamma'(t_0) \neq \hat{0}$, decimos que la recta definida (paramétricamente) como*

$$R(t) := \gamma(t_0) + t\gamma'(t_0)$$

es la recta tangente a C en $\hat{x}_0$.

Con relación a esta última definión, es necesario hacer algunas observasiones importantes. La primera de ellas es que, aun cuando el "aspecto" geométrico del conjunto C en el punto $\hat{x}_0$ nos de la impresión de que existe la recta tangente, esto no significa que con cualquier parametrización de C sea posible "calcularla". Esto se debe fundamentalmente al hecho de que, a pesar de que exista una parametrización $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ de C tal que $\gamma(t_0) = \hat{x}_0$ y γ sea derivable en t_0, nada nos garantiza que se satisfaga la condición de que $\gamma'(t_0) \neq \hat{0}$. En el siguiente ejemplo mostramos un conjunto y dos parametrizaciones de éste que ilustran las dos posibilidades en el cálculo de la recta tangente.

Ejemplo 3.1 *Sea $C = \{(x, y) \in \mathbb{R}^2 \mid y = x^2\}$. Considere las siguientes parametrizaciones de C.*

1. *Sea $\gamma : \mathbb{R} \to \mathbb{R}^2$ dada por*

$$\gamma(t) = (t, t^2).$$

Como el lector fácilmente puede comprobar, γ es una parametriazación de C. En particular en el punto $\gamma(0) = (0, 0)$ se tiene que

$$\begin{aligned}\gamma'(0) &= \lim_{t\to 0}\frac{\gamma(t)-\gamma(0)}{t-0}\\ &= \lim_{t\to 0}(1,t)\\ &= (1,0),\end{aligned}$$

de tal forma que la recta

$$\begin{aligned}l(t) &= \gamma(0)+t\gamma'(0)\\ &= (0,0)+t(1,0)\\ &= (t,0),\end{aligned}$$

que no es más que el eje X, es tangente a C en el punto $(0,0)$.

2. *Sea $\tilde{\gamma}:\mathbb{R}\to\mathbb{R}^2$ dada por*

$$\tilde{\gamma}(t) = (t^3, t^6).$$

Como en el caso anterior, también es fácil comprobar que $\tilde{\gamma}$ es una parametrización de C y que $\tilde{\gamma}(0)=(0,0)$. Por otra parte, se tiene que

$$\begin{aligned}\lim_{t\to 0}\frac{\tilde{\gamma}(t)-\tilde{\gamma}(0)}{t-0} &= \lim_{t\to 0}(t^2,t^5)\\ &= (0,0),\end{aligned}$$

de modo que $\tilde{\gamma}'(0)=(0,0)$ y por lo tanto la parametrización $\tilde{\gamma}$ no proporciona un vector que permita construir la recta tangente a C en el $(0,0)$.

Otra observación importante es que la existencia misma de $\gamma'(t_0)$ no tiene mucho que ver con el "aspecto geométrico" que tenga el conjunto C en el punto $\hat{x}_0=\gamma(t_0)$. A diferencia de lo que sucedía con el "aspecto geométrico" de la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$ cuando ésta era derivable en un punto, si el conjunto C tiene un "pico" en $\hat{x}_0$, esto no significa que

$\gamma'(t_0)$ no exista, como el lector puede comprobar fácilmente en el caso de la cicloide (figura 3.3), la cual tiene un "pico" en el punto $(\pi, 0) = \gamma(\pi)$ a pesar de que $\gamma'(\pi)$ sí existe (lo que se deduce de un resultado que probaremos más adelante).

Otra ilustración de este mismo hecho lo tenemos con la gráfica de la función valor absoluto, como lo veremos en el siguiente

Ejemplo 3.2 *Sea $C \subset \mathbb{R}^2$ la gráfica de la función $f(x) = |x|$; es decir, $C = \{(x, |x|) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$. Considere las siguientes parametrizaciones de C.*

1. *Sea $\gamma : \mathbb{R} \to \mathbb{R}^2$ dada por*

$$\gamma(t) = (t, |t|).$$

Nótese que para esta parametrización de C (la más "común"), no existe $\gamma'(0)$ pues por una parte

$$\begin{aligned} \lim_{t \to 0^+} \frac{\gamma(t) - \gamma(0)}{t - 0} &= \lim_{t \to 0^+} \frac{1}{t}(t, t) \\ &= (1, 1), \end{aligned}$$

mientras que

$$\begin{aligned} \lim_{t \to 0^-} \frac{\gamma(t) - \gamma(0)}{t - 0} &= \lim_{t \to 0^-} \frac{1}{t}(t, -t) \\ &= (1, -1), \end{aligned}$$

de tal forma que, por el problema 31 del capítulo 2,

$$\gamma'(0) = \lim_{t \to 0} \frac{\gamma(t) - \gamma(0)}{t - 0}$$

no existe.

2. *Sea ahora $\tilde{\gamma} : \mathbb{R} \to \mathbb{R}^2$ dada por*

$$\tilde{\gamma}(t) = \begin{cases} (-t^2, t^2) & \text{si } t \leq 0 \\ \\ (t^2, t^2) & \text{si } t \geq 0 \end{cases}$$

También es fácil ver que esta es una parametrización de C. Nótese que ahora

$$\begin{aligned}\lim_{t\to 0^+}\frac{\tilde{\gamma}(t)-\tilde{\gamma}(0)}{t-0} &= \lim_{t\to 0^+}\frac{1}{t}(t^2,t^2)\\ &= \lim_{t\to 0^+}(t,t)\\ &= (0,0)\end{aligned}$$

y

$$\begin{aligned}\lim_{t\to 0^-}\frac{\tilde{\gamma}(t)-\tilde{\gamma}(0)}{t-0} &= \lim_{t\to 0^-}\frac{1}{t}(-t^2,t^2)\\ &= \lim_{t\to 0^-}(-t,t)\\ &= (0,0),\end{aligned}$$

de tal forma que, por el mismo problema 31 del capítulo 2, se tiene que $\tilde{\gamma}'(0)$ sí existe y es igual al vector $(0,0)$.

Con base en los ejemplos anteriores se concluye que, si $C \subset \mathbb{R}^2$ es un conjunto que se puede parametrizar por una función γ y $\gamma'(t)$ existe, si este vector es diferente del vector $\hat{0}$, sólo entonces se podrá asegurar que C es un conjunto que no tiene un "pico" en el punto $\gamma(t)$ (en cuyo caso, por razones obvias, diremos que el conjunto C es "suave" en el punto $\gamma(t)$). Y si $\gamma'(t) = \hat{0}$, entonces no podemos asegurar nada sobre el aspecto geométrico de C en el punto $\gamma(t)$.

Para concluir con las observaciones relacionadas con la definición de recta tangente que hemos dado, es importante recordar que este es un concepto que el lector ya conoce para cierto tipo de conjuntos en $\mathbb{R}^2$, específicamente, aquellos que se obtienen como la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$.

Por otra parte, como vimos en la primera sección de este capítulo, si $f : I \subset \mathbb{R} \to \mathbb{R}$, sabemos que la gráfica de f es un subconjunto de $\mathbb{R}^2$ definido como

$$G_f := \big\{(x, f(x)) \in \mathbb{R}^2 \mid x \in I\big\}$$

y que la función $\gamma : I \subset \mathbb{R} \to \mathbb{R}^2$ dada por $\gamma(t) = (t, f(t))$ es una parametrización de G_f.

En el problema 8 de este capítulo, el lector probará que, si la función f es derivable en $t_0 \in I$, entonces la parametrización γ también es derivable en t_0 y además, que la recta tangente dada en la definición 3.3 no es más que una forma paramétrica de la recta tangente que le fue definida en su primer curso de cálculo diferencial.

De la observación anterior se concluye que la definición de recta tangente dada en la defínción 3.3 "amplía" a la familia de subconjuntos de $\mathbb{R}^2$ para los cuales ahora tenemos definido este concepto.

Lo que haremos a continuación será dar un ejemplo de un subconjunto de $\mathbb{R}^2$ que no se puede ver (completo) como la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$ (aunque en el capítulo 5 probaremos un importante resultado que nos asegura que algunas partes de éste sí se pueden ver de esa forma), pero que sí se puede parametrizar. Por lo tanto, con base en la definición 3.3, podremos calcular su recta tangente en algunos de sus puntos.

Adicional a lo anterior, en este ejemplo también mostraremos cómo se puede encontrar la parametrización de un subconjunto de $\mathbb{R}^2$ que está definido a través de coordenadas polares.

Ejemplo 3.3 *Sea* $C = \{\hat{x} = (r, \theta) \in \mathbb{R}^2 \mid 0 \leq \theta \leq 2\pi, r = 2(1 + \cos(\theta))\}$, *en donde la pareja* (r, θ) *representa coordenadas polares del punto* $\hat{x}$. *Este conjunto es conocido como la curva* cardioide *(ver figura 3.5).*

Lo primero que haremos será calcular una parametrización de C. *Dado que este conjunto está descrito en términos de coordenadas polares, y que las operaciones de suma y producto por un escalar de vectores (operaciones que es necesario realizar para el cálculo de la derivada de una parametrización) no se pueden expresar en forma sencilla en este tipo de coordenadas, lo primero que haremos será describir a los elementos de* C *en términos de coordenadas cartesianas.*

De esta forma, aplicando las ecuaciones de cambio de coordenadas vistas en el capítulo 1, tendremos que, si (x, y) *representan las coordenadas cartesianas de* $\hat{x} \in C$, *entonces*

$$\begin{aligned} x &= r\cos(\theta) \\ &= 2(1+\cos(\theta))\cos(\theta) \end{aligned}$$

y

$$\begin{aligned} y &= r\,\mathrm{sen}(\theta) \\ &= 2(1+\cos(\theta))\,\mathrm{sen}(\theta). \end{aligned}$$

Es decir,

$$C = \left\{(2(1+\cos(\theta))\cos(\theta), 2(1+\cos(\theta))\,\mathrm{sen}(\theta)) \in \mathbb{R}^2 \mid 0 \leq \theta \leq 2\pi\right\},$$

de modo que si definimos $\gamma : [0, 2\pi] \to \mathbb{R}^2$ *como*

$$\gamma(\theta) = (2(1+\cos(\theta))\cos(\theta), 2(1+\cos(\theta))\,\mathrm{sen}(\theta)),$$

γ *será una parametrización de* C *expresada en términos de las coordenadas cartesianas de sus elementos.*

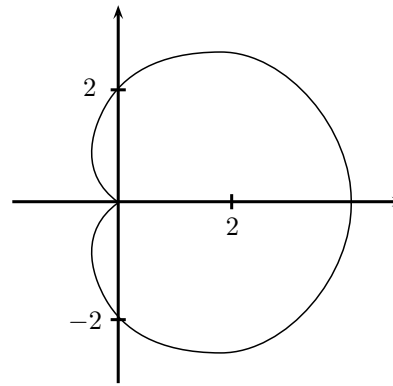


Figura 3.5: La curva cardioide cuya ecuación polar está dada por $r = 2(1 + cos(\theta))$.

Ahora que ya contamos con una parametrización de C, para calcular la derivada de ésta, haremos uso del primer resultado que formularemos en la siguiente sección (proposición 3.1), lo que sin lugar a dudas constituye un pequeño y breve abuso, que esperamos el lector disculpe.

Con base en este resultado, tendremos que

$$\begin{aligned}\gamma'(\theta) &= (-2\operatorname{sen}(\theta)\cos(\theta) - 2(1+\cos(\theta))\operatorname{sen}(\theta),\\ &\qquad -2\operatorname{sen}(\theta)\operatorname{sen}(\theta) + 2(1+\cos(\theta))\cos(\theta))\\ &= \left(-2\operatorname{sen}(\theta) - 4\operatorname{sen}(\theta)\cos(\theta), 2\cos(\theta) + 2\left(\cos^2(\theta) - \operatorname{sen}^2(\theta)\right)\right)\\ &= \left(-2\operatorname{sen}(\theta) - 2\operatorname{sen}(2\theta), 2\cos(\theta) + 2\cos(2\theta)\right),\end{aligned}$$

de donde podemos concluir que, por ejemplo, la recta tangente a la cardioide en el punto $\gamma(\pi/2) = (0,2)$ es la recta parametrizada por

$$\begin{aligned}\gamma(\pi/2) + t\gamma'(\pi/2) &= (0,2) + t\,(-2,-2)\\ &= (-2t, -2t+2)\,,\end{aligned}$$

es decir, la recta cuya ecuación cartesiana es $y = x + 2$. Análogamente, la recta tangente a la cardioide en el punto $\gamma(3\pi/2) = (0,-2)$ es la recta parametrizada por

$$\begin{aligned}\gamma(3\pi/2) + t\gamma'(3\pi/2) &= (0,-2) + t\,(2,-2)\\ &= (-2t, 2t-2)\,,\end{aligned}$$

es decir, la recta cuya ecuación cartesiana es $y = -x - 2$.

Concluimos esta sección mencionando otra interpretación del concepto de derivada que hemos introducido.

Supongamos que utilizamos una función $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ para describir el movimiento de un objeto. En este caso, el parámetro (o la variable) t del cual depende la función γ representará al tiempo, de tal forma que para un valor específico $t_0 \in I$, $\gamma(t_0)$ representará su posición en este instante. Sin duda el lector estará de acuerdo en que si, $t \in I$ es otro instante muy cercano a t_0, entonces el vector

$$\frac{1}{t - t_0}(\gamma(t) - \gamma(t_0)) = \frac{\gamma(t) - \gamma(t_0)}{t - t_0}$$

contiene información muy valiosa del movimiento descrito por γ cerca del instante t_0.

En efecto, este vector nos da información de hacia "dónde" se está moviendo nuestro objeto (a partir de su posición en el tiempo t_0), y su norma nos da un promedio de la "rapidez" con la que se está moviendo cerca del instante t_0.

En virtud de lo anterior, se suele decir que este vector representa la "*velocidad promedio*" de nuestro objeto durante el intervalo de tiempo comprendido entre t_0 y t.

Como seguramente el lector ya está imaginando, si

$$\lim_{t\to t_0} \frac{\gamma(t)-\gamma(t_0)}{t-t_0}$$

existe, dicho valor límite (que también es un vector) se puede interpretar como la velocidad de nuestro objeto en el instante t_0, o como se suele decir, *la velocidad instantánea* de nuestro objeto, en el instante t_0.

3.3. Propiedades de la derivada

Como hicimos en el caso de los conceptos de límite y continuidad, lo siguiente que haremos será mostrar la estrecha relación que existe entre la derivada de una función $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ y sus funciones coordenadas $\gamma_1, \ldots, \gamma_n$ (en un sistema de referencia dado), las cuales serán funciones de $\mathbb{R}$ en $\mathbb{R}$ y para las que ya conocemos el concepto de derivada.

Como seguramente el lector ya intuye, la derivabilidad de la función γ es una condición necesaria y suficiente de la derivabilidad de sus funciones coordenadas, resultado que dejamos plasmado en la siguiente

Proposición 3.1 *Sea $\gamma = (\gamma_1, \ldots, \gamma_n) : I \subset \mathbb{R} \to \mathbb{R}^n$ y $t_0 \in I$. La función γ es derivable en t_0 si y sólo si cada función coordenada γ_i es derivable en t_0, para $i \in \{1, \ldots, n\}$. En ambos casos se tiene que*

$$\gamma'(t_0) = (\gamma_1'(t_0), \ldots, \gamma_n'(t_0)).$$

La prueba de esta proposición es una consecuencia inmediata de la proposición 2.7 del capítulo 2 y se deja al lector.

Lo siguiente que haremos será mostrar la relación que existe entre el concepto de derivada y las operaciones aritméticas entre funciones de este tipo. Con base en la proposición anterior, los correspondientes resultados de derivación para funciones de $\mathbb{R}$ en $\mathbb{R}$ y la proposición 2.9 del capítulo 2, la prueba de estas propiedades es inmediata y también se deja al lector.

Proposición 3.2 *Sean $\gamma, \tilde{\gamma} : I \subset \mathbb{R} \to \mathbb{R}^n$, $t_0 \in I$ y $h : I \subset \mathbb{R} \to \mathbb{R}$. Si las funciones γ, $\tilde{\gamma}$ y h son derivables en t_0, entonces:*

1. *$\gamma + \tilde{\gamma}$ es derivable en t_0 y además*

$$(\gamma + \tilde{\gamma})'(t_0) = \gamma'(t_0) + \tilde{\gamma}'(t_0)$$

2. *$h\gamma$ es derivable en t_0 y además*

$$(h\gamma)'(t_0) = h'(t_0)\gamma(t_0) + h(t_0)\gamma'(t_0)$$

   En particular, si h es la función constante c, entonces

$$(c\gamma)'(t_0) = c\gamma'(t_0)$$

3. *$\gamma \cdot \tilde{\gamma} : I \subset \mathbb{R} \to \mathbb{R}$ es derivable en t_0 y además*

$$(\gamma \cdot \tilde{\gamma})'(t_0) = \gamma'(t_0) \cdot \tilde{\gamma}(t_0) + \gamma(t_0) \cdot \tilde{\gamma}'(t_0)$$

4. *si $n = 3$, $\gamma \times \tilde{\gamma}$ es derivable en t_0 y además*

$$(\gamma \times \tilde{\gamma})'(t_0) = \gamma'(t_0) \times \tilde{\gamma}(t_0) + \gamma(t_0) \times \tilde{\gamma}'(t_0).$$

Otras operaciones que podemos realizar con una función de $\mathbb{R}$ en $\mathbb{R}^n$, son la composición por la izquierda con una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ y por la derecha con una función de $\mathbb{R}^m$ en $\mathbb{R}$.

Dado que habrá que esperar hasta los capítulos 3 y 4 para contar con un concepto de derivada para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ (con $n > 1$) y que hasta

este momento sólo sabemos cómo derivar funciones de $\mathbb{R}^m$ en $\mathbb{R}$ cuando $m = 1$, por ahora nos vamos a conformar con formular un resultado (¡la primera *regla de la cadena* de este texto!) que establece condiciones para que la composición (por la derecha) de una función de $\mathbb{R}$ en $\mathbb{R}^n$ con otra de $\mathbb{R}$ en $\mathbb{R}$ sea derivable, y una fórmula para calcular su derivada.

Proposición 3.3 *Sean $J \subset \mathbb{R}$ un intervalo, $h : J \subset \mathbb{R} \to \mathbb{R}$, $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$, $t_0 \in I$ y $x_0 \in J$ tales que $h(J) \subset I$ y $h(x_0) = t_0$. Si h es derivable en x_0 y γ es derivable en t_0, entonces $\gamma \circ h$ es derivable en x_0 y además*

$$\begin{aligned}(\gamma \circ h)'(x_0) &= \gamma'(h(x_0))h'(x_0)\\ &= \gamma'(t_0)h'(x_0).\end{aligned}$$

Si $\gamma = (\gamma_1, \ldots, \gamma_n)$, entonces $\gamma \circ h = (\gamma_1 \circ h, \ldots, \gamma_n \circ h)$, de modo que la prueba de esta proposición también es una consecuencia inmediata de la proposición 3.1 y de la regla de la cadena para funciones de $\mathbb{R}$ en $\mathbb{R}$, por lo que se deja al lector.

La prueba también se puede hacer sin recurrir a las funciones coordenadas y a la proposición 3.1, imitando la prueba para funciones de $\mathbb{R}$ en $\mathbb{R}$, lo que se pide hacer en el problema 6 de este capítulo.

En cuanto a la fórmula para calcular $(\gamma \circ h)'(x_0)$, es importante mencionar que, aun cuando para escribir la multiplicación de un escalar por un vector siempre hemos puesto primero al escalar, en esta ocasión lo hemos escrito después del vector con la idea de conservar la forma en que se suele expresar la regla de la cadena.

Como en el caso de las funciones de $\mathbb{R}$ en $\mathbb{R}$, la derivabilidad de una función de $\mathbb{R}$ en $\mathbb{R}^n$ en un punto t_0 implica la continuidad de ésta en dicho punto. Como es de esperarse, esto también es una consecuencia inmediata de la proposición 3.1 y del correspondiente resultado para el caso real, lo que dejamos expresado en la siguiente

Proposición 3.4 *Sean $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ y $t_0 \in I$. Si γ es derivable en t_0, entonces γ es continua en t_0.*

Aun cuando ya mencionamos que la prueba de esta proposición es una consecuencia inmediata de la proposición 3.1 y del correspondiente resultado para el caso real, ésta también se puede hacer imitando la prueba que se hace en este caso observando que

$$\gamma(t) - \gamma(t_0) = (t - t_0)\frac{\gamma(t) - \gamma(t_0)}{t - t_0}$$

para todo $t \neq t_0$, de modo que

$$\begin{aligned}\lim_{t\to t_0}(\gamma(t) - \gamma(t_0)) &= \lim_{t\to t_0}(t - t_0)\frac{\gamma(t) - \gamma(t_0)}{t - t_0}\\ &= (0)\gamma'(t_0)\\ &= \hat{0},\end{aligned}$$

de donde concluimos que γ es continua en t_0.

Sin duda, el resultado más importante relacionado con el concepto de derivada de funciones de $\mathbb{R}$ en $\mathbb{R}$, es el Teorema del Valor Medio. Lamentablemente este teorema no se puede generalizar a funciones de $\mathbb{R}$ en $\mathbb{R}^n$ y los siguientes ejemplos ilustran tan desafortunado hecho.

Ejemplo 3.4 *Considere las siguientes funciones:*

1. $\gamma : [-1, 1] \subset \mathbb{R} \to \mathbb{R}^2$ *dada por*

$$\gamma(t) = \begin{cases} (-t^2, t^2) & si\ -1 \leq t \leq 0 \\ (t^2, t^2) & si\ 0 \leq t \leq 1 \end{cases}$$

Nótese que

$$\gamma'(t) = \begin{cases} 2t(-1, 1) & si\ -1 \leq t < 0 \\ (0, 0) & si\ t = 0 \\ 2t(1, 1) & si\ 0 < t \leq 1 \end{cases}$$

de modo que no existe $\xi \in (-1,1)$ *para el cual se satisfaga que*

$$\begin{aligned}\gamma(1)-\gamma(-1) &= (1-(-1))\gamma'(\xi)\\ &= 2\gamma'(\xi),\end{aligned}$$

puesto que el vector $\gamma(1)-\gamma(-1)=(1,1)-(-1,1)=(2,0)$ *no es un múltiplo escalar de* $\gamma'(t)$ *para ningún valor de* $t \in (-1,1)$ *(figura 3.6* (a)*).*

2. $\tilde{\gamma} : [0,1] \subset \mathbb{R} \to \mathbb{R}^3$ *dada por* $\tilde{\gamma}(t) = (\cos(2\pi t), \operatorname{sen}(2\pi t), t)$*. Entonces*

$$\tilde{\gamma}'(t) = (-2\pi \operatorname{sen}(2\pi t), 2\pi \cos(2\pi t), 1),$$

de modo que no existe $\xi \in (0,1)$ *tal que*

$$\begin{aligned}\tilde{\gamma}(1)-\tilde{\gamma}(0) &= (2\pi-0)\tilde{\gamma}'(\xi)\\ &= (-4\pi \operatorname{sen}(\xi), 4\pi\cos(\xi), 2\pi),\end{aligned}$$

puesto que $\tilde{\gamma}(1)-\tilde{\gamma}(0)=(0,0,1)$ *y* $2\pi \neq 1$*. De hecho, dado que para ningún valor de* t *las funciones* $\operatorname{sen}(2\pi t)$ *y* $\cos(2\pi t)$ *son simultáneamente cero (recuerde que* $\operatorname{sen}^2(2\pi t)+\cos^2(2\pi t)=1$ *para toda* $t \in \mathbb{R}$*), el vector* $\tilde{\gamma}(1)-\tilde{\gamma}(0)$ *nunca es "paralelo" (o múltiplo escalar) de* $\tilde{\gamma}'(t)$ *para ningún valor de* t *(figura 3.6* (b)*).*

No obstante los ejemplos anteriores, se puede formular una cierta "versión" parecida al Teorema de Valor Medio (sólo para funciones de $\mathbb{R}$ en $\mathbb{R}^2$) que se puede probar justamente como una consecuencia del Teorema del Valor Medio Generalizado de Cauchy para funciones de $\mathbb{R}$ en $\mathbb{R}$, y que se deja como problema al lector (problema 12).

Concluiremos esta sección introduciendo el concepto de derivada de orden superior para una función de $\mathbb{R}$ en $\mathbb{R}^n$. De forma análoga a lo que sucede en el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$, si una función $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ es derivable para cada t del intervalo $J \subset I$, entonces la función $\gamma' : J \subset \mathbb{R} \to \mathbb{R}^n$ que a cada $t \in J$ asocia la derivada de γ en t, es decir $t \mapsto \gamma'(t)$ es nuevamente una función de $J \subset \mathbb{R}$ en $\mathbb{R}^n$.

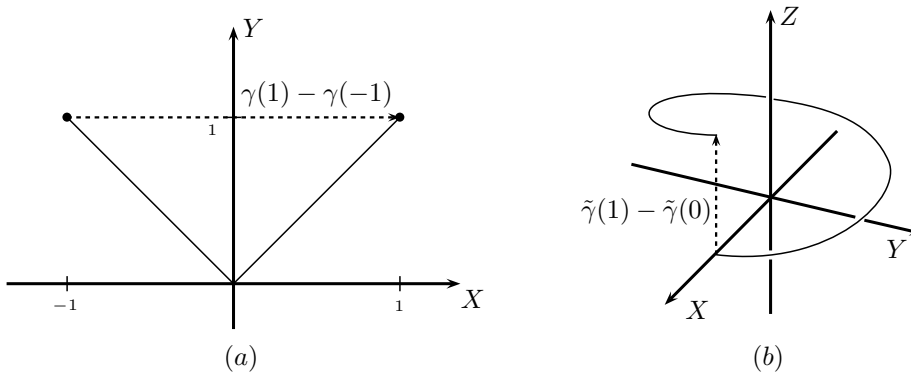


Figura 3.6: El teorema del Valor Medio no se puede generalizar a funciones de $\mathbb{R}$ en $\mathbb{R}^n$. Como lo muestran estos ejemplos, los vectores $\gamma(1)-\gamma(-1)$ y $\tilde{\gamma}(1)-\tilde{\gamma}(0)$ no son paralelos a la derivada de γ y $\tilde{\gamma}$, respectivamente, evaluada en algún punto intermedio.

Si esta función también resulta ser derivable en cada punto de J, se define entonces la segunda derivada de γ, que denotamos por γ'' o por $\gamma^{(2)}$, como la función $\gamma'' : J \subset \mathbb{R} \to \mathbb{R}^n$ dada por

$$\begin{aligned}\gamma''(t) &:= (\gamma')'(t)\\ &:= \lim_{h\to 0} \frac{\gamma'(t+h)-\gamma'(t)}{h}\end{aligned}$$

para cada $t \in J$.

Siguiendo este procedimiento, en general podemos dar de manera inductiva la siguiente

Definición 3.4 *Sean $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ derivable para cada t del intervalo $J \subset I$, y $n \in \mathbb{N}$. Si $\gamma^{(1)}(t) := \gamma'(t)$ y $\gamma^{(n)}$ está definida y es derivable para cada $t \in J$, definimos (inductivamente) la $n+1$ derivada de γ en t, que denotamos por $\gamma^{(n+1)}(t)$, como*

$$\begin{aligned}\gamma^{(n+1)}(t) &:= (\gamma^{(n)})'(t) \\ &:= \lim_{h\to 0} \frac{\gamma^{(n)}(t+h) - \gamma^{(n)}(t)}{h}\end{aligned}$$

para cada $t \in J$.

En las siguientes secciones daremos las posibles interpretaciones que se le pueden dar a las derivadas de orden superior, dependiendo del contexto en el que se esté trabajando.

3.4. Derivada y geometría

Una vez que ya contamos con el concepto de derivada para funciones de $\mathbb{R}$ en $\mathbb{R}^n$, tenemos todo lo necesario para definir el concepto de *curva*, el cual constituye uno de los usos más importantes de este tipo de funciones.

Definición 3.5 *Sea* $C \subset \mathbb{R}^n$. *Decimos que* C *es una curva si existe* $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ *derivable en el intervalo* I *tal que* $\gamma(I) = C$. *Si además* $\gamma'(t) \neq \hat{0}$ *para toda* $t \in I$ *decimos que* C *es una curva suave (o regular). Como ya se mencionó en la definición 3.1, decimos que* γ *es una parametrización de* C.

Si γ es una función derivable que parametriza a un conjunto C, además de proporcionarnos una forma de encontrar una ecuación (parámetrica) de su recta tangente en $\gamma(t)$ (si $\gamma'(t) \neq \hat{0}$), la derivada de γ nos puede ser útil para "resolver" el siguiente problema geométrico: apoyados en la idea intuitiva que tengamos de lo que es (¡o debiera de ser!) la "longitud" de C, ¿cómo calcular ésta "longitud"?

Para abordar este problema, vamos a suponer que la curva se puede parametrizar por una función $\gamma : [a, b] \subset \mathbb{R} \to \mathbb{R}^n$, la cual es derivable en el intervalo $[a, b]$ y que además es inyectiva.

Es intuitivamente claro que si elegimos un número finito de puntos en C, digamos $\hat{x}_0, \ldots, \hat{x}_k$, de tal forma que $\hat{x}_0$ y que $\hat{x}_k$ sean sus "extremos" y que $\hat{x}_{i-1}$ y $\hat{x}_i$ sean "consecutivos" para $i \in \{1, \ldots, k\}$, entonces la suma

$$\sum_{i=1}^{k} \|\hat{x}_i - \hat{x}_{i-1}\|$$

es una aproximación a la "longitud" de C, y que esta aproximación será mejor en la medida de que tomemos más (y muy bien distribuidos) puntos en C (ver figura 3.7).

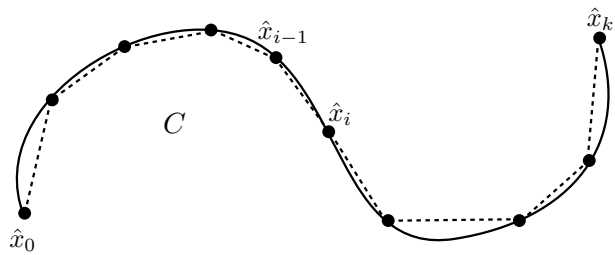


Figura 3.7: La suma de las distancias entre los puntos consecutivos $\hat{x}_{i-1}$ y $\hat{x}_i$ es una muy buena aproximación a la "longitud" de C.

De la suposición de que γ parametriza a la curva C y que γ es inyectiva, al conjunto de puntos $\{\hat{x}_0, \ldots, \hat{x}_k\}$ le corresponde una partición $\mathcal{P} = \{t_0, \ldots, t_k\}$ del intervalo $[a, b]$ tal que $\hat{x}_i = \gamma(t_i)$ para $i \in \{1, \ldots, k\}$. De esta forma, tenemos entonces que

$$\sum_{i=1}^{k} \|\hat{x}_i - \hat{x}_{i-1}\| = \sum_{i=1}^{k} \|\gamma(t_i) - \gamma(t_{i-1})\|$$

y la suma de la derecha será una buena aproximación a la "longitud" de C si $\mathcal{P}$ es una partición muy "fina" del intervalo $[a, b]$. Ahora lo que haremos será analizar más de cerca cómo se puede expresar cada vector $\gamma(t_i) - \gamma(t_{i-1})$.

Aun cuando ya mostramos que el Teorema del Valor Medio no se cumple en general para funciones de $\mathbb{R}$ en $\mathbb{R}^n$, lo que sí podemos asegurar es que, si $\gamma = (\gamma_1, \ldots, \gamma_n)$, aplicando el Teorema del Valor Medio (para funciones de $\mathbb{R}$ en $\mathbb{R}$) a cada una de estas funciones coordenadas, entonces para cada $j \in \{1, \ldots, n\}$ se sabe que existe $\xi_i^{(j)} \in (t_{i-1}, t_i)$ tal que

$$\gamma_j(t_i) - \gamma_j(t_{i-1}) = \gamma_j'\left(\xi_i^{(j)}\right)(t_i - t_{i-1})$$

de tal forma que

$$\begin{aligned}\gamma(t_i) - \gamma(t_{i-1}) &= (\gamma_1(t_i) - \gamma_1(t_{i-1}), \ldots, \gamma_n(t_i) - \gamma_n(t_{i-1})) \\ &= \left(\gamma_1'\left(\xi_i^{(1)}\right)(t_i - t_{i-1}), \ldots, \gamma_n'\left(\xi_i^{(n)}\right)(t_i - t_{i-1})\right) \\ &= \left(\gamma_1'\left(\xi_i^{(1)}\right), \ldots, \gamma_n'\left(\xi_i^{(n)}\right)\right)(t_i - t_{i-1}).\end{aligned}$$

Si ahora suponemos también que cada función coordenada γ_j' es continua (lo que es equivalente a que γ' sea continua) en el intervalo cerrado $[a, b]$, entonces el vector

$$\left(\gamma_1'\left(\xi_i^{(1)}\right), \ldots, \gamma_n'\left(\xi_i^{(n)}\right)\right)$$

se "parecerá" a $\gamma'(\xi_i)$ para cualquier $\xi_i \in (t_{i-1}, t_i)$ (y este "parecido" será mucho mejor nuevamente cuando la partición $\mathcal{P}$ sea una partición muy "fina" del intervalo $[a, b]$) de tal forma que

$$\begin{aligned}\|\gamma(t_i) - \gamma(t_{i-1})\| &= \left\|\left(\gamma_1'\left(\xi_i^{(1)}\right), \ldots, \gamma_n'\left(\xi_i^{(n)}\right)\right)(t_i - t_{i-1})\right\| \\ &\approx \left\|\gamma'(\xi_i)(t_i - t_{i-1})\right\| \\ &= \left\|\gamma'(\xi_i)\right\|(t_i - t_{i-1}).\end{aligned}$$

En virtud de lo anterior, si la suma

$$\sum_{i=1}^{k} \|\hat{x}_i - \hat{x}_{i-1}\|$$

es una buena "aproximación" a la longitud de C en la medida de que la partición $\mathcal{P}$ sea una partición cada vez más "fina" del intervalo $[a, b]$, entonces con la suma

$$\sum_{i=1}^{k} \left\|\gamma'(\xi_i)\right\|(t_i - t_{i-1})$$

sucederá lo mismo, y lo relevante de la conclusión anterior es que esta última suma tiene la pecularidad de ser una suma de Riemann correspondiente a la función $f(t) = \|\gamma'(t)\|$ (para $t \in [a, b]$), por lo que dicha suma se "aproxima" a la integral

$$\int_a^b \left\|\gamma'(t)\right\| dt \tag{3.4}$$

y esta aproximación será mejor justo cuando ¡$\mathcal{P}$ sea una partición muy "fina" del intervalo $[a, b]$!

Resumiendo la discusión anterior (¡y todas las aproximaciones que ahí aparecen!), todo parece indicar que, si $C \subset \mathbb{R}^n$ es una curva que se puede parametrizar por una función $\gamma : [a, b] \subset \mathbb{R} \to \mathbb{R}^n$, la cual es derivable en el intervalo $[a, b]$, inyectiva y además $\gamma' : [a, b] \subset \mathbb{R} \to \mathbb{R}^n$ es continua, entonces su "longitud" deberá estar dada por la integral 3.4. Con el fin de "reforzar" esta conclusión, mostraremos en el siguiente ejemplo que esto sucede así, calculando la integral anterior con una curva (y una parametrización de ella) para la cual ya conocemos su longitud.

Ejemplo 3.5 *Sea $C \subset \mathbb{R}^2$ la circunferencia de radio $r > 0$ con centro en un punto $\hat{x}_0 = (x_0, y_0)$.*

Mostraremos que, si tomamos la parametrización de C dada por la función

$$\gamma(t) = (r\cos(t) + x_0, r\,\mathrm{sen}(t) + y_0)$$

con $t \in [0, 2\pi]$, y calculamos la integral dada en 3.4, dicha integral vale $2\pi r$, que como sabemos desde hace mucho tiempo, es el perímetro de la cincunferencia C.

En efecto, nótese que $\gamma'(t) = (-r\,\mathrm{sen}(t), r\cos(t))$ de modo que

$$\left\|\gamma'(t)\right\| = r$$

para toda $t \in [0, 2\pi]$ y por lo tanto tenemos que

$$\int_a^b \left\|\gamma'(t)\right\| dt = \int_0^{2\pi} r dt$$
$$= 2\pi r.$$

Como mencionamos al inicio de toda esta discusión, la realidad es que a estas alturas no contamos (salvo por algunos casos específicos) con una definición de "longitud" de una curva, y justo lo que estamos a punto de hacer es llenar ese vacío.

En realidad definiremos (aprovechando que la integral 3.4 sólo depende de la función γ) lo que llamaremos la "longitud asociada a esta parametrización". Cuando γ cumpla con ciertas condiciones (como por ejemplo, que sea inyectiva, salvo tal vez por un número finito de puntos de su dominio) podremos interpretar dicha "longitud asociada" como "*la* longitud" de la curva que está parametrizada por γ.

Todo ello lo recogemos en la siguiente

Definición 3.6 *Sea $\gamma : [a,b] \subset \mathbb{R} \to \mathbb{R}^n$ tal que $\|\gamma'(t)\|$ es integrable en $[a,b]$. Definimos la longitud de γ (que denotamos por $l(\gamma)$) como*

$$l(\gamma) := \int_a^b \left\|\gamma'(t)\right\| dt.$$

Si γ es una función inyectiva en $[a,b]$, salvo quizás por un número finito de puntos, y $C = \gamma([a,b])$, entonces el número $l(\gamma)$ dado en la definición anterior se puede interpretar como "la longitud" de C. De hecho, con base en esta idea de "longitud", podemos definir una función $\alpha : [a,b] \subset \mathbb{R} \to \mathbb{R}$ como

$$\alpha(t) = \int_a^t \left\|\gamma'(u)\right\| du,$$

cuyo valor en $t \in [a,b]$ también se puede interpretar como "la longitud del subarco" de C dado por $\gamma([a,t])$.

Si además de la inyectividad de γ en "casi" todo su dominio, tenemos que $\|\gamma'(t)\|$ es una función continua de t (lo que se satisface si γ' es continua en $[a,b]$), por el Teorema Fundamental del Cálculo sabemos que α será una función derivable y que

$$\alpha'(t) = \left\|\gamma'(t)\right\| \geq 0,$$

lo que significa que α siempre resulta ser una función no decreciente en el intervalo $[a,b]$ cuya imagen es el intervalo $[0,l(\gamma)]$.

Lo interesante de esta función α es que, si además $\gamma'(t) \neq \hat{0}$ para toda $t \in [a,b]$ (es decir que $C = \gamma([a,b])$ es una curva suave (o regular)), entonces

$$\alpha'(t) = \left\|\gamma'(t)\right\| > 0,$$

de modo que en este caso α es una función estrictamente creciente en el intervalo $[a,b]$, y por tanto su función inversa $\alpha^{-1} : [0,l(\gamma)] \subset \mathbb{R} \to [a,b] \subset \mathbb{R}$ existe, la cual, dado que α' es continua y $\alpha'(t) \neq 0$ para toda $t \in [a,b]$, por el Teorema de la Función Inversa (para funciones de $\mathbb{R}$ en $\mathbb{R}$), también es derivable.

Lo importante de la discusión anterior es que, si $C = \gamma([a,b])$ es una curva suave (o regular), entonces a través de la función α^{-1} podemos obtener otra parametrización de C, dada por

$$\tilde{\gamma} = \gamma \circ \alpha^{-1} : [0,l(\gamma)] \subset \mathbb{R} \to \mathbb{R}^n,$$

la cual resulta tener propiedades muy relevantes.

A este tipo de parametrización de C se le conoce con el nombre de *parametrización por longitud de arco* (para el caso de la parametrización $\tilde{\gamma}$ este nombre es muy adecuado puesto que su parámetro $s = \alpha(t) \in [0,l(\gamma)]$ es justo la longitud del subarco $\gamma([a,t])$ de C), y la última parte de esta sección la dedicaremos a mostrar algunas de sus características.

Antes de hacer esto, es importante llamar la atención sobre la forma en que construimos la parametrización $\tilde{\gamma}$ a partir de la parametrización original γ, que es un caso particular de una relación más general que puede haber entre dos parametrizaciones de una misma curva C. Esta relación da

lugar al concepto de *reparametrización*, el cual definiremos antes de estudiar las propiedades específicas de la parametrización (o reparametrización) por longitud de arco.

Definición 3.7 *Sean $C \subset \mathbb{R}^n$ una curva y $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ una parametrización (derivable) de C. Decimos que $\tilde{\gamma} : J \subset \mathbb{R} \to \mathbb{R}^n$, una función derivable, es una reparametrización de γ si existe $\alpha : J \subset \mathbb{R} \to \mathbb{R}$ derivable, tal que $\alpha(J) = I$ y $\tilde{\gamma}(s) = (\gamma \circ \alpha)(s)$ para toda $s \in J$. Si $\alpha'(s) \geq 0$ para toda $s \in J$, decimos que $\tilde{\gamma}$ preserva la orientación de γ, y si $\alpha'(s) \leq 0$ para toda $s \in J$, decimos que $\tilde{\gamma}$ invierte la orientación de γ.*

Dada una parametrización $\gamma : [a, b] \subset \mathbb{R} \to \mathbb{R}^n$ de una curva $C \subset \mathbb{R}^n$, una reparametrización de γ que siempre se puede construir es la que se obtiene a partir de la función $\alpha : [a, b] \to [a, b]$ dada por $\alpha(t) = a + b - t$. Como $\alpha(a) = b$ y $\alpha(b) = a$, la función $\gamma \circ \alpha$ tiene la propiedad de que $(\gamma \circ \alpha)(a) = \gamma(b)$ y $(\gamma \circ \alpha)(b) = \gamma(a)$, es decir, $\gamma \circ \alpha$ parametriza al conjunto C con la particularidad de que su punto inicial $(\gamma \circ \alpha)(a)$ es el punto final $\gamma(b)$ de γ, y su punto final $(\gamma \circ \alpha)(b)$ es el punto inicial $\gamma(a)$ de γ (ver figura 3.8), razón por la cual a esta reparametrización se le suele denotar por $-\gamma$.

Por todo lo anterior, y dado que

$$\begin{aligned}(-\gamma)'(t) &= (\gamma \circ \alpha)'(t) \\ &= \gamma'(\alpha(t))\alpha'(t) \\ &= -\gamma'(a + b - t),\end{aligned}$$

se tiene que $-\gamma$ recorre a C en la dirección contraria en que la recorre γ.

Como se mostró en los párrafos previos a la definición 3.7, y a diferencia de la reparametrización $-\gamma$, la reparametrización por longitud de arco $\tilde{\gamma}$ sólo se puede obtener si $C \subset \mathbb{R}^n$ es una curva suave y $\gamma : [a, b] \subset \mathbb{R} \to \mathbb{R}^n$ es una parametrización de C tal que $\gamma'(t)$ es continua y diferente de $\hat{0}$ para toda $t \in [a, b]$.

Sin duda una propiedad muy característica de la reparametrización por longitud de arco $\tilde{\gamma}$ de una curva suave C, es la referente a la rapidez con que la "recorre". En efecto, como se puede observar en la reparametrización

$\tilde{\gamma}$ que se obtiene a partir de una parametrización γ definida sobre un intervalo $[a, b]$, $\tilde{\gamma}$ queda definida sobre el intervalo $[0, l(\gamma)]$ de tal manera que su parámetro s "varía" dentro de un intervalo que tiene la misma longitud que la curva C. De esta forma, pareciera que para "recorrer" a esta curva, a la parametrización $\tilde{\gamma}$ le bastaría con "recorrer" una "unidad de distancia" sobre C por cada "unidad de distancia" que "recorra" el parámetro s, lo que en términos de rapidez, significa que ésta última tendría que valer 1. Esta propiedad, que efectivamente tiene la reparametrización $\tilde{\gamma}$, servirá más adelante para definir lo que significa en general que una curva $C \subset \mathbb{R}^n$ esté parametrizada por longitud de arco.

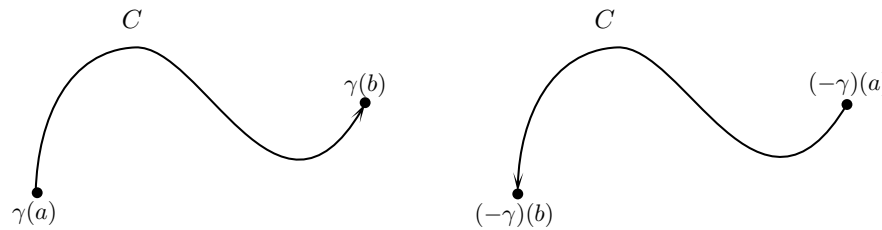


Figura 3.8: Las parametrizaciones γ y $-\gamma$ de la misma curva C.

Definición 3.8 *Sea $C \subset \mathbb{R}^n$ una curva y $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ una parametrización (derivable) de C. Decimos que γ es una parametrización por longitud de arco de C si $\|\gamma'(t)\| = 1$ para toda $t \in I$.*

De la definición anterior se desprende inmediatamente que si $C \subset \mathbb{R}^n$ es una curva que posee una parametrización por longitud de arco, entonces C es una curva suave (o regular).

Lo que a continuación haremos será mostrar que una curva suave (o regular) $C \subset \mathbb{R}^n$ que tiene una parametrización $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ que además de ser derivable, tiene derivada continua, tiene una parametrización por longitud de arco. La forma en que construiremos esta parametrización es básicamente la misma que usamos para contruir la reparametrización $\tilde{\gamma}$ en los párrafos anteriores.

Proposición 3.5 *Sea $C \subset \mathbb{R}^n$ una curva suave. Si $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ es una parametrización de C tal que $\gamma'(t)$ es continua y diferente de $\hat{0}$ para toda $t \in I$, entonces C tiene una parametrización por longitud de arco.*

Demostración. Sea $t_0 \in I$ un punto fijo; definimos $\alpha : I \subset \mathbb{R} \to \mathbb{R}$ como

$$\alpha(t) = \int_{t_0}^{t} \left\|\gamma'(u)\right\| du. \tag{3.5}$$

Por el Teorema Fundamental del Cálculo sabemos que α es derivable y que

$$\alpha'(t) = \left\|\gamma'(t)\right\|$$

para toda $t \in I$. De esta forma, dado que $\alpha'(t) = \|\gamma'(t)\| > 0$ (además de ser continua), por el Teorema de la Función Inversa (para funciones de $\mathbb{R}$ en $\mathbb{R}$), sabemos que α es invertible y que su inversa α^{-1}, que está definida sobre el intervalo $\alpha(I)$, también es derivable.

Definimos $\tilde{\gamma} : \alpha(I) \subset \mathbb{R} \to \mathbb{R}^n$ como $\tilde{\gamma}(s) = (\gamma \circ \alpha^{-1})(s)$ para cada $s \in \alpha(I)$. Claramente $\tilde{\gamma}$ es una parametrización de C (puesto que es una reparametrización de γ), y por la regla de la cadena probada en la proposición 3.3, tenemos que

$$\begin{aligned}\tilde{\gamma}'(s) &= (\gamma \circ \alpha^{-1})'(s) \\ &= \gamma'(\alpha^{-1}(s))(\alpha^{-1})'(s)\end{aligned}$$

de modo que

$$\begin{aligned}\tilde{\gamma}'(s) &= \gamma'(\alpha^{-1}(s))(\alpha^{-1})'(s) \\ &= \gamma'(\alpha^{-1}(s))\frac{1}{\alpha'(\alpha^{-1}(s))} \\ &= \gamma'(\alpha^{-1}(s))\frac{1}{\|\gamma'(\alpha^{-1}(s))\|},\end{aligned}$$

de donde concluimos que

$$\left\|\tilde{\gamma}'(s)\right\| = 1$$

para cada $s \in \alpha(I)$. ■

Con relación a la construcción de la parametrización por longitud de arco $\tilde{\gamma}$ que se da en la prueba de la proposición anterior, es importante hacer la siguiente observación.

Desde un punto de vista teórico, la construcción es impecable; sin embargo (y para ser sinceros), desde un punto de vista práctico, obtener una expresión explícita para $\tilde{\gamma}$ puede resultar ser una tarea muy difícil (y en algunos casos ¡imposible!). Esta dificultad (o imposibilidad) está relacionada con el hecho de que en la construcción de $\tilde{\gamma}$ está involucrada la obtención de la inversa de la función α dada por la identidad 3.5, lo que no siempre se puede hacer de manera explícita.

Para no contribuir con una posible causa de depresión de los amables lectores de este texto, daremos un ejemplo en el que sí es posible calcular $\tilde{\gamma}$.

Ejemplo 3.6 *Sea $C \subset \mathbb{R}^3$ la curva parametrizada por la función $\gamma : \mathbb{R} \to \mathbb{R}^3$ dada por*

$$\gamma(t) = (\cos(t), \operatorname{sen}(t), t)$$

(ver figura 3.9).

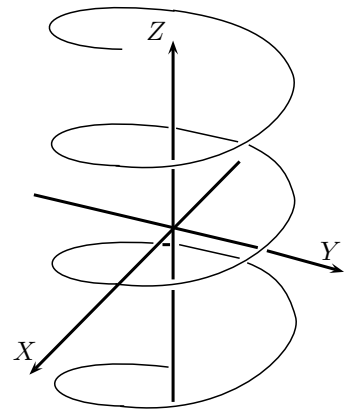


Figura 3.9: La curva (hélice) del ejemplo 3.6.

Calcularemos, con base en γ, una parametrización por longitud de arco de C. Definimos $\alpha : \mathbb{R} \to \mathbb{R}$ como

$$\begin{aligned}\alpha(t) &= \int_0^t \left\|\gamma'(u)\right\| du \\ &= \int_0^t \sqrt{2} du \\ &= \sqrt{2}t.\end{aligned}$$

Entonces, despejando la variable t de la ecuación

$$\begin{aligned}s &= \alpha(t) \\ &= \sqrt{2}t\end{aligned}$$

obtenemos que

$$\begin{aligned}t &= \frac{s}{\sqrt{2}} \\ &= \alpha^{-1}(s),\end{aligned}$$

de tal forma que

$$\begin{aligned}\tilde{\gamma}(s) &= (\gamma \circ \alpha^{-1})(s) \\ &= \gamma(\alpha^{-1}(s)) \\ &= \gamma(s/\sqrt{2}) \\ &= (\cos(s/\sqrt{2}), \operatorname{sen}(s/\sqrt{2}), s/\sqrt{2})\end{aligned}$$

para $s \in \mathbb{R}$.

Una vez que ya hemos definido lo que es una parametrización por longitud de arco (y mostrado cómo y cuándo podemos obtener una de éstas), lo que haremos a continuación será mostrar la forma en que se puede usar para definir algunas características geométricas de las curvas.

Dado que la característica principal de una parametrización por longitud de arco γ de una curva $C \subset \mathbb{R}^n$, es que su rapidez es constante 1 (es decir, $\|\gamma'(s)\| = 1$), si podemos calcular la derivada de γ' (es decir γ''), la información que ésta nos proporciona sólo está relacionada con el cambio de dirección de γ' o con la forma en que se "dobla" la curva C.

De manera más precisa, y dado que por el problema 10 ya sabemos que $\gamma''(s)$ siempre es perpendicular a $\gamma'(s)$, la manera en que (valga la redundancia) se "*curva*" la curva C, deberá estar contenida en la norma de $\gamma''(s)$. Que este número tiene esta información puede comprobarse calculándolo para una circunferencia de radio r, en cuyo caso, además de depender de r, deberá ser el mismo en cualquier punto de ésta.

Ejemplo 3.7 *Sea $C \subset \mathbb{R}^2$ la circunferencia de radio $r > 0$ con centro en el origen y $\gamma : \mathbb{R} \to \mathbb{R}^2$ una parametrización por longitud de arco de C dada por*

$$\gamma(s) = \left(r \cos\left(\frac{s}{r}\right), r \operatorname{sen}\left(\frac{s}{r}\right) \right).$$

Entonces tenemos que

$$\gamma'(s) = \left(-\operatorname{sen}\left(\frac{s}{r}\right), \cos\left(\frac{s}{r}\right) \right)$$

y

$$\gamma''(s) = \left(-\frac{1}{r}\cos\left(\frac{s}{r}\right), -\frac{1}{r}\operatorname{sen}\left(\frac{s}{r}\right) \right),$$

de modo que

$$\left\|\gamma''(s)\right\| = \frac{1}{r}$$

para toda $s \in \mathbb{R}$.

Además de confirmar nuestras sospechas, el ejemplo anterior ilustra un hecho geométrico muy intuitivo: si el radio r de una circunferencia es muy grande, ésta está menos curvada, y si el radio es muy pequeño, entonces la circunferencia está muy curvada. Ambos hechos se ven reflejados en el número $\|\gamma''(s)\|$, puesto que en el primer caso éste es más cercano a 0, mientras que en el segundo caso es muy grande. Con base en estos razonamientos y en el ejemplo anterior, damos la siguiente

Definición 3.9 *Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^n$ tal que $\gamma''(s)$ existe para cada $s \in I$. Definimos la curvatura de C en el punto $\gamma(s) \in C$, que denotamos por $k(s)$, como*

$$k(s) := \left\|\gamma''(s)\right\|.$$

Como ya habíamos mencionado, si γ es una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^n$, los vectores $\gamma'(s)$ y $\gamma''(s)$ son perpendiculares, y dado que $\gamma'(s)$ es tangente a C (además de tener norma 1), entonces se dice que $\gamma''(s)$ "apunta" en la dirección normal a la curva C (que es la dirección en la que ésta se "curva"). Con base en estas observaciones, establecemos las siguientes definiciones y notaciones.

Definición 3.10 *Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^n$ tal que $\gamma''(s)$ existe para toda $s \in I$.*

1. *Definimos el vector tangente unitario a C en $\gamma(s)$ (determinado por γ), el cual denotamos por $T(s)$, como*

$$T(s) := \gamma'(s).$$

2. *Si $\gamma''(s) \neq \hat{0}$, definimos el vector normal unitario a C en $\gamma(s)$, el cual denotamos por $N(s)$, como*

$$N(s) := \frac{\gamma''(s)}{\|\gamma''(s)\|}.$$

3. *Si $\gamma''(s) \neq \hat{0}$, definimos el plano osculador a C en $\gamma(s)$, el cual denotamos por Π_s, como el plano generado por lo vectores $T(s)$ y $N(s)$ que contiene al punto $\gamma(s)$. Es decir*

$$\Pi_s := \{\alpha T(s) + \beta N(s) + \gamma(s) \mid \alpha, \beta \in \mathbb{R}\}.$$

4. *Si $\gamma''(s) \neq \hat{0}$, definimos la circunferencia osculadora a C en $\gamma(s)$, a la cual denotamos por C_s, como la circunferencia contenida en el plano osculador con centro en el punto*

$$\gamma(s) + \frac{1}{k(s)} N(s).$$

llamado centro de curvatura, y radio

$$\frac{1}{k(s)},$$

A este último número lo llamaremos el radio de curvatura de la curva C en $\gamma(s)$.

5. *Si $C \subset \mathbb{R}^3$ y $\gamma''(s) \neq \hat{0}$, definimos el vector binormal unitario a C en $\gamma(s)$, al cual denotamos por $B(s)$, como*

$$B(s) := T(s) \times N(s).$$

No podemos dejar de mencionar que en las dos definiciones anteriores haría falta mostrar que los conceptos ahí plasmados (salvo por los vectores tangente y binormal unitarios), son todos independientes de la parametrización por longitud de arco que se esté usando. Aunque probar esto no es muy difícil, hacerlo sí resultaría bastante laborioso y por lo mismo escapa a los objetivos de este texto.

Para ilustrar todos los "objetos" asociados a una curva suave dados en la definición anterior, damos el siguiente

Ejemplo 3.8 *Sea $C \subset \mathbb{R}^3$ la "hélice" del ejemplo 3.6 y su parametrización por longitud de arco*

$$\gamma(s) = (\cos(s/\sqrt{2}), \operatorname{sen}(s/\sqrt{2}), s/\sqrt{2})$$

que ahí calculamos.

Tenemos entonces que:

1. *el vector tangente unitario $T(s)$ está dado por*

$$\begin{aligned} T(s) &= \gamma'(s) \\ &= \left(-\frac{1}{\sqrt{2}} \operatorname{sen}(s/\sqrt{2}), \frac{1}{\sqrt{2}} \cos(s/\sqrt{2}), \frac{1}{\sqrt{2}} \right) \end{aligned}$$

2. la curvatura $k(s)$ está dada por

$$\begin{aligned} k(s) &= \|\gamma''(s)\| \\ &= \left\| \left(-\frac{1}{2}\cos(s/\sqrt{2}), -\frac{1}{2}\operatorname{sen}(s/\sqrt{2}), 0 \right) \right\| \\ &= \frac{1}{2} \end{aligned}$$

3. el vector normal unitario $N(s)$ está dado por

$$\begin{aligned} N(s) &= \frac{\gamma''(s)}{\|\gamma''(s)\|} \\ &= \left(-\cos(s/\sqrt{2}), -\operatorname{sen}(s/\sqrt{2}), 0 \right) \end{aligned}$$

4. el plano osculador Π_s, expresado en forma "paramétrica", está dado

$$\begin{aligned} \Pi_s &= \{\alpha T(s) + \beta N(s) + \gamma(s) \mid \alpha, \beta \in \mathbb{R}\} \qquad (3.6) \\ &= \Big\{ 1/\sqrt{2} \Big(-\alpha \operatorname{sen}(s/\sqrt{2}) + \sqrt{2}(1-\beta)\cos(s/\sqrt{2}), \\ &\qquad (\alpha \cos(s/\sqrt{2}) + \sqrt{2}(1-\beta)\operatorname{sen}(s/\sqrt{2}), \alpha + s \Big) \mid \alpha, \beta \in \mathbb{R} \Big\}. \end{aligned}$$

Como Π_s está generado por los vectores $T(s)$ y $N(s)$, aprovechando que estamos en $\mathbb{R}^3$, podemos recurrir al vector binormal para obtener un vector normal a Π_s y con éste calcular su ecuación cartesiana.

De esta forma, dado que

$$\begin{aligned} B(s) &= T(s) \times N(s) \\ &= \frac{1}{\sqrt{2}} \left(\operatorname{sen}(s/\sqrt{2}), -\cos(s/\sqrt{2}), 1 \right) \end{aligned}$$

y que Π_s contiene al punto $\gamma(s)$, su ecuación cartesiana estará dada por

$$\begin{aligned} \left(\operatorname{sen}(s/\sqrt{2}), -\cos(s/\sqrt{2}), 1 \right) \cdot ((x,y,z) - \\ (\cos(s/\sqrt{2}), \operatorname{sen}(s/\sqrt{2}), s/\sqrt{2})) = 0, \end{aligned}$$

la cual queda expresada como

$$\operatorname{sen}(s/\sqrt{2})x - \cos(s/\sqrt{2})y + z = \frac{s}{\sqrt{2}}$$

y que, como es de esperarse, es satisfecha por las ternas dadas en la identidad 3.6.

5. *finalmente, el centro de la circunferencia osculadora C_s está dado por*

$$\begin{aligned}\gamma(s) + \frac{1}{k(s)}N(s) &= (\cos(s/\sqrt{2}), \operatorname{sen}(s/\sqrt{2}), s/\sqrt{2}) \\ &\quad + \frac{1}{1/2}\left(-\cos(s/\sqrt{2}), -\operatorname{sen}(s/\sqrt{2}), 0\right) \\ &= \left(-\cos(s/\sqrt{2}), -\operatorname{sen}(s/\sqrt{2}), s/\sqrt{2}\right),\end{aligned}$$

de modo que, como dicha circunferencia está contenida en el plano generado por los vectores $T(s)$ y $N(s)$, y su radio es $1/k(s) = 2$, una expresión paramétrica para C_s estaría dada por

$$\gamma(s) + 2N(s) + 2\cos(\theta)T(s) + 2\operatorname{sen}(\theta)N(s)$$

con $\theta \in [0, 2\pi]$.

En la figura 3.10 se muestran algunos de los "objetos" que calculamos en el ejemplo anterior, incluyendo la curva determinada por los centros de curvatura de C, la cual recibe el nombre de *evoluta* de la curva C y que en este caso es nuevamente una hélice.

Una primera consecuencia interesante (e inmediata) de las definiciones 3.9 y 3.10, es la relación existente entre los vectores $T'(s)$ y $N(s)$, la cual dejamos plasmada en la siguiente proposición y cuya prueba dejamos al lector.

Proposición 3.6 *Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ una parametrización por longitud de arco de una curva suave $C \subset \mathbb{R}^n$. Si $\gamma''(s)$ existe y es diferente de $\hat{0}$, entonces*

$$T'(s) = k(s)N(s).$$

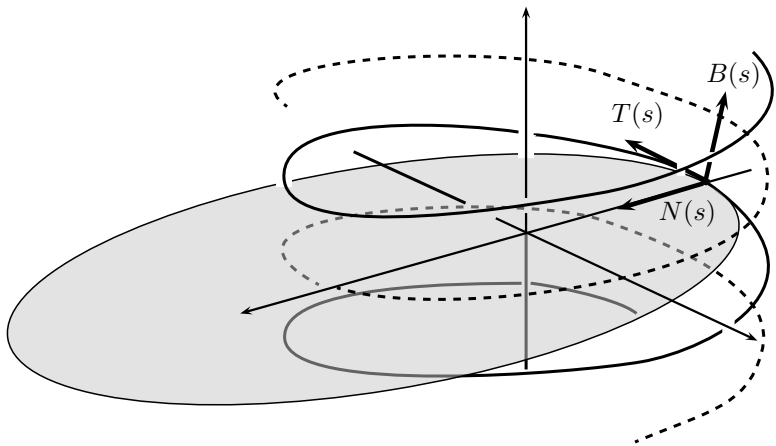


Figura 3.10: La curva hélice y algunos de sus "objetos" calculados en el ejemplo 3.8, incluyendo su evoluta.

De forma análoga a la proposición anterior, cuando nuestra curva C está contenida en $\mathbb{R}^3$ se tiene que el vector $B'(s)$ es un múltiplo escalar del vector $N(s)$, afirmación que probamos en la siguiente proposición y con base en la cual introducimos el concepto de *torsión* para curvas suaves contenidas en $\mathbb{R}^3$.

Proposición 3.7 *Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una parametrización por longitud de arco de una curva suave $C \subset \mathbb{R}^3$. Si para $s \in I$ se tiene que $B'(s)$ existe, entonces $B'(s)$ es un múltiplo escalar de $N(s)$.*

Demostración. Dado que

$$\begin{aligned} \|B(s)\| &= \|T(s) \times N(s)\| \\ &= \|T(s)\| \, \|N(s)\| \operatorname{sen}(\pi/2) \\ &= 1, \end{aligned}$$

por el problema 10 de este capítulo sabemos que $B'(s)$ es perpendicular a $B(s)$, es decir $B(s) \cdot B'(s) = 0$.

Por otra parte, de la definición de $B(s)$, el inciso 4 de la proposición 3.2 y la identidad de la proposición 3.6, tenemos que

$$\begin{aligned} B'(s) &= T'(s) \times N(s) + T(s) \times N'(s) \\ &= k(s)(N(s) \times N(s)) + T(s) \times N'(s) \\ &= T(s) \times N'(s), \end{aligned}$$

de modo que $B'(s)$ también es perpendicular a $T(s)$, es decir $T(s) \cdot B'(s) = 0$.

Por tanto, dado que la terna $\{T(s), N(s), B(s)\}$ es una base (ortonormal) de $\mathbb{R}^3$, entonces existen $\alpha, \beta, \gamma \in \mathbb{R}$ tales que

$$B'(s) = \alpha T(s) + \beta N(s) + \gamma B(s),$$

de donde obtenemos que

$$\begin{aligned} 0 &= T(s) \cdot B'(s) \\ &= T(s) \cdot (\alpha T(s) + \beta N(s) + \gamma B(s)) \\ &= \alpha \end{aligned}$$

y

$$\begin{aligned} 0 &= B(s) \cdot B'(s) \\ &= B(s) \cdot (\alpha T(s) + \beta N(s) + \gamma B(s)) \\ &= \gamma \end{aligned}$$

y por lo tanto, que

$$B'(s) = \beta N(s),$$

que es lo deseábamos demostrar. ■

Como habíamos mencionado, con base en la afirmación de la proposición anterior definimos el concepto de *torsión* para curvas suaves contenidas en $\mathbb{R}^3$, que desde un punto de vista geométrico, se interpreta como una medida de "qué tanto se tuerce una curva" (y consecuentemente "qué tanto tiende a salirse de un plano", a diferencia de lo que sucede con la curvatura, que puede "curvarse" sin salirse de éste).

Definición 3.11 *Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una parametrización por longitud de arco de una curva suave $C \subset \mathbb{R}^3$. Si para $s \in I$ se tiene que $B'(s)$ existe, definimos la torsión de la curva C en el punto $\gamma(s)$, que denotamos por $\tau(s)$, como el número real tal que*

$$B'(s) = -\tau(s)N(s).$$

El uso del signo negativo en la definición anterior no tiene ningún significado específico y obedece sólo a razones de caracter histórico (al parecer, ¡así se usó desde que se introdujo!).

Para ilustrar este concepto, calcularemos la torsión de la hélice del ejemplo 3.8 aprovechando los cálculos que ahí mismo hicimos.

Ejemplo 3.9 *Sea $C \subset \mathbb{R}^3$ la "hélice" del ejemplo 3.8 y su parametrización por longitud de arco*

$$\gamma(s) = (\cos(s/\sqrt{2}), \operatorname{sen}(s/\sqrt{2}), s/\sqrt{2}).$$

De ese mismo ejemplo sabemos que

$$N(s) = \left(-\cos(s/\sqrt{2}), -\operatorname{sen}(s/\sqrt{2}), 0\right)$$

y

$$B(s) = \frac{1}{\sqrt{2}}\left(\operatorname{sen}(s/\sqrt{2}), -\cos(s/\sqrt{2}), 1\right),$$

de modo que

$$B'(s) = \frac{1}{2}\left(\cos(s/\sqrt{2}), \operatorname{sen}(s/\sqrt{2}), 0\right)$$

y por tanto, tenemos que

$$\tau(s) = \frac{1}{2}$$

para toda $s \in \mathbb{R}$.

Para concluir esta sección y completar las expresiones dadas en la proposición 3.6 y la definición 3.11, daremos una expresión del vector $N'(s)$

para el caso de curvas suaves en $\mathbb{R}^3$. Como hemos venido mencionando, los vectores $\{T(s), N(s), B(s)\}$ forman una base (ortonormal) de $\mathbb{R}^3$ y por esta simple razón el vector $N'(s)$ siempre se puede expresar como una combinación lineal de éstos.

Lo más interesante es que en realidad el vector $N'(s)$ sólo se escribe como combinación lineal de los vectores $T(s)$ y $B(s)$, y que los coeficientes de esta combinación lineal están dados por $-k(s)$ y $\tau(s)$, respectivamente. Este hecho es lo que probaremos en la siguiente

Proposición 3.8 *Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una parametrización por longitud de arco de una curva suave $C \subset \mathbb{R}^3$. Si para $s \in I$ se tiene que $N'(s)$ existe, entonces*

$$N'(s) = -k(s)T(s) + \tau(s)B(s).$$

Demostración. De la definición del vector $B(s)$, y de acuerdo con la "regla de la mano derecha" del producto vectorial, sabemos que

$$N(s) = B(s) \times T(s)$$

y por la regla de derivación del producto vectorial (inciso (4) de la proposición 3.2), tenemos que

$$N'(s) = B'(s) \times T(s) + B(s) \times T'(s),$$

de modo que, sustituyendo las identidades de la proposición 3.6 y la definición 3.11, obtenemos que

$$\begin{aligned} N'(s) &= B'(s) \times T(s) + B(s) \times T'(s) \\ &= (-\tau(s)N(s)) \times T(s) + B(s) \times (k(s)N(s)) \\ &= -k(s)(N(s) \times B(s)) + \tau(s)(T(s) \times N(s)) \\ &= -k(s)T(s) + \tau(s)B(s). \end{aligned}$$

■

Las identidades

$$\begin{aligned} T'(s) &= k(s)N(s) \\ N'(s) &= -k(s)T(s) + \tau(s)B(s) \\ B'(s) &= -\tau(s)N(s) \end{aligned}$$

son conocidas como las *Fórmulas de Frenet-Serret*[4] y tienen importantes aplicaciones en la cinemática.

3.5. Derivada y movimiento

Para concluir este capítulo, mostraremos algunas conclusiones importantes que se pueden obtener cuando a una función $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ se le interpreta como una forma de describir el movimiento de un objeto. Como ya habíamos mencionado anteriormente, bajo esta interpretación se tiene entonces que $\gamma'(t)$ y $\gamma''(t)$ representan la velocidad y la aceleración de dicho objeto, y $r(t) = \|\gamma'(t)\|$ su rapidez, todas ellas al tiempo t.

Un primer resultado importante que abordaremos es aquél que nos dice cuáles son las "componentes" de la aceleración en dos direcciones básicas del movimiento descrito por γ: la dirección "tangente" y la dirección "normal" a este movimiento.

Como vimos en las secciones anteriores, la dirección "tangente" nos la proporciona el vector $\gamma'(t)$ de tal forma que, si hacemos (bajo el supuesto de que $\gamma'(t) \neq \hat{0}$ para toda $t \in I$)

[4] Jean Frédéric Frenet (Périgueux, Francia, 7 de febrero de 1816 - Périgueux, Francia, 12 de junio de 1900), fue un famoso matemático francés que introdujo la Teoría de Curvas junto a Joseph Serret.

Joseph Alfred Serret (París, Francia, 30 de agosto de 1819 - Versalles, Francia, 2 de marzo de 1885), más conocido como Joseph Serret, fue un matemático francés famoso por desarrollar junto a Jean Frenet la Teoría de Curvas.

(Fuente: Wikipedia)

$$T(t) = \frac{\gamma'(t)}{\|\gamma'(t)\|} = \frac{\gamma'(t)}{r(t)}, \tag{3.7}$$

entonces $T(t)$ es un vector en la dirección "tangente" cuya norma es constante uno, y por el problema 10 tendremos que $T'(t)$ es perpendicular a $T(t)$ de modo que $T'(t)$ nos proporcionará la dirección "normal" al movimiento.

De esta forma, si $T'(t) \neq \hat{0}$, hacemos

$$N(t) = \frac{T'(t)}{\|T'(t)\|}$$

y bajo todos los supuestos anteriores, lo que mostraremos en la siguiente proposición será la forma explícita en que $\gamma''(t)$ se expresa como combinación (lineal) de los vectores $T(t)$ y $N(t)$.

Proposición 3.9 *Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ tal que $T(t)$ existe para toda $t \in I$. Si $T'(t) \neq \hat{0}$, entonces*

$$\gamma''(t) = r'(t)T(t) + r^2(t)k(t)N(t), \tag{3.8}$$

en donde $k(t)$ es la curvatura (de la curva descrita por γ) en el punto $\gamma(t)$.

Demostración. Por la identidad 3.7, tenemos que

$$\gamma'(t) = r(t)T(t),$$

de tal forma que, usando la regla de derivación del inciso 2 de la proposición 3.2, se tiene que

$$\begin{aligned}\gamma''(t) &= r'(t)T(t) + r(t)T'(t) \\ &= r'(t)T(t) + r(t)\left\|T'(t)\right\| \frac{T'(t)}{\|T'(t)\|} \\ &= r'(t)T(t) + r(t)\left\|T'(t)\right\| N(t)\end{aligned}$$

y por el inciso (a) del problema 29, tenemos que

$$\gamma''(t) = r'(t)T(t) + r^2(t)k(t)N(t).$$

■

La identidad 3.8 amerita algunos comentarios. El primero de ellos es que si la rapidez de un objeto es constante, entonces $r'(t) = 0$ y por lo tanto la magnitud de la componente tangencial de su aceleración es cero, lo que significa que ésta última actúa sólo en la dirección normal al movimiento, hecho que ya habíamos mencionado cuando tratamos con parametrizaciones por longitud de arco y el tema de la curvatura.

El segundo comentario es acerca de la magnitud de la componente normal de la aceleración, la cual es igual al cuadrado de la rapidez del objeto, multiplicado por la curvatura de la curva que éste describe. Seguramente el lector ya se ha dado cuenta de que esto explica por qué un automóvil se vuelca cuando se toma una curva muy "cerrada" (curvatura grande) a una rapidez también muy grande.

Lo siguiente que haremos será mostrar un par de propiedades que deben satisfacer aquellos objetos (como por ejemplo un planeta o un cometa) que se mueven bajo la influencia de la gravedad de un objeto de masa muy grande (como por ejemplo una estrella). Para ello, primero recordaremos la *Ley de la Gravitación Universal* formulada por Newton, en la que se establece cuál es la fuerza ejercida entre dos cuerpos de masas m_1 y m_2 separados por una distancia r.

Ley de la Gravitación Universal

La magnitud F de la fuerza ejercida entre dos cuerpos de masas m_1 y m_2 que están separados por una distancia r es directamente proporcional al producto de sus masas e inversamente proporcional al cuadrado de su distancia, es decir, que

$$F = G\frac{m_1 m_2}{r^2}.$$

La fuerza $\hat{F}$ (cuya magnitud es F) ejercida entre ambos cuerpos actúa en la dirección de la línea que los une. El número G es conocido como la constante de la Gravitación Universal.

Nótese que, si el centro de masa de cada uno de los cuerpos de masa m_1 y m_2 está ubicado en el punto $\hat{x}_1$ y $\hat{x}_2$ respectivamente, y suponemos que $m_2 > m_1$, entonces la fuerza de atracción $\hat{F}$ que ejerce el objeto de masa m_2 sobre el objeto de masa m_1 estará dada por

$$\hat{F} = G\frac{m_1 m_2}{\|\hat{x}_2 - \hat{x}_1\|^3}(\hat{x}_2 - \hat{x}_1). \tag{3.9}$$

Otra ley de la física de la que también echaremos mano, es la *Segunda ley del movimiento* (igualmente formulada por Newton) la cual establece que, si un objeto de masa m se mueve debido a la acción de un campo de fuerzas (en el caso que analizaremos, el campo de fuerzas gravitatorias determinadas por una estrella (o sol) de masa M), entonces la fuerza ejercida sobre el objeto (en la posición que éste tenga en cada instante t de su movimiento) también se puede obtener multiplicando la masa del objeto por su aceleración en dicho instante. Es decir, si $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ es la función que nos asocia la posición del objeto (de su centro de masa, para ser más exactos) en cada instante t, y denotamos por $\hat{F}(\gamma(t))$ la fuerza ejercida sobre éste en la posición $\gamma(t)$, entonces

$$\begin{aligned}\hat{F}(\gamma(t)) &= ma(t)\\ &= m\gamma''(t).\end{aligned}$$

Lo que probaremos en la siguiente proposición será que, si un objeto se mueve bajo la acción de un campo de fuerzas como el que describimos anteriormente, entonces dicho movimiento necesariamente se tiene que llevar a cabo sobre un plano (fijo). Este hecho es importante para probar la *Primera Ley de Kepler*, la cual establece que la órbita (o curva) descrita por el objeto es una elipse[5].

[5]Para los interesados en la prueba de las leyes de Kepler, se puede consultar en [3].

Proposición 3.10 *Si un objeto de masa m se mueve debido a la acción de un campo de fuerzas gravitatorias (determinadas por otro objeto de masa M que está fijo), entonces el movimiento del primer objeto se lleva a cabo sobre un plano fijo (al que también pertenece el objeto de masa M).*

Demostración. Fijemos un sistema cartesiano cuyo origen se encuentre en el centro de masa del objeto de masa M, y sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ la función que nos asocia la posición del centro de masa del objeto de masa m en cada instante t. Mostraremos que el producto vectorial

$$\gamma(t) \times \gamma'(t)$$

no depende de t, es decir, que es un vector constante $\hat{c}$, de donde el movimiento descrito por γ se llevará a cabo en el plano que es perpendicular a este vector. Para ello, derivaremos este producto y mostraremos que dicha derivada es el vector $\hat{0}$ para toda $t \in I$.

Primero notemos que, por la Ley de Gravitación Universal y de acuerdo con la identidad 3.9, dado que el origen de nuestro sistema coordenado está ubicado en el centro de masa del segundo objeto, la fuerza ejercida por éste en cada posición $\gamma(t)$ está dada por

$$\hat{F}(\gamma(t)) = -G\frac{mM}{\|\gamma(t)\|^3}\gamma(t).$$

Por otra parte, y de acuerdo con la segunda ley del movimiento, tenemos que

$$\hat{F}(\gamma(t)) = m\gamma''(t),$$

de tal forma que de estas dos identidades, concluimos que $\gamma(t)$ y $\gamma''(t)$ son vectores paralelos y por tanto que

$$\gamma(t) \times \gamma''(t) = \hat{0}$$

para toda $t \in I$.

Ahora, derivando el producto vectorial $(\gamma \times \gamma')(t) = \gamma(t) \times \gamma'(t)$, se tiene que

$$\begin{aligned}(\gamma \times \gamma')'(t) &= \gamma'(t) \times \gamma'(t) + \gamma(t) \times \gamma''(t) \\ &= \hat{0}\end{aligned}$$

lo que significa que

$$\gamma(t) \times \gamma'(t) = \hat{c} \tag{3.10}$$

para toda $t \in I$. ■

Para concluir esta breve sección, y como una continuación del análisis del movimiento de nuestro objeto de masa m (debido a la acción del campo gravitatorio de un segundo objeto de masa M), desarrollaremos todo lo indispensable para probar lo que se conoce como la *Segunda Ley de Kepler*. Para ello, y dando por cierto que dicho movimiento se realiza en un plano, vamos a establecer un sistema cartesiano XYZ de tal forma que la curva descrita por el objeto esté contenida en el plano XY. Daremos también como un hecho que la trayectoria seguida es una curva cerrada que "rodea" al origen (de hecho, una elipse, como mencionamos párrafos arriba).

Lo importante de los supuestos anteriores es que, si recurrimos a las coordenadas polares (ρ, θ) para describir la posición del objeto, y escribimos a su coordenada ρ como una función del ángulo θ (es decir que $\rho = f(\theta)$), entonces el área A encerrada por la curva descrita y dos semirectas con ángulos θ_1 y θ_2 $(\theta_1 < \theta_2)$ (ver figura 3.11), está dada por la expresión

$$A = \frac{1}{2} \int_{\theta_1}^{\theta_2} f^2(\theta) d\theta.$$

Con base en lo anterior, podemos formular la *Segunda Ley de Kepler* en los siguientes términos: si $\theta(t)$ es la función que nos da el ángulo θ del vector de posición al tiempo t, y

$$A(\theta) = \frac{1}{2} \int_{\theta_0}^{\theta} f^2(u) du \tag{3.11}$$

es la función del ángulo θ que nos da el área "barrida" a partir de un ángulo fijo θ_0, entonces la función $(A\circ\theta)(t) = A(\theta(t))$ tiene razón de cambio constante, es decir, que $(A \circ \theta)'(t)$ es constante. Dicho de manera menos técnica, esto significa que "el vector de posición de nuestro objeto "barre" áreas iguales en tiempos iguales".

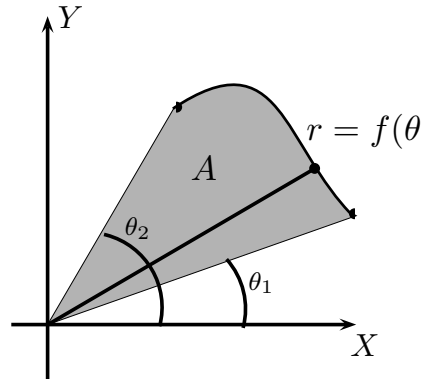


Figura 3.11: La región A encerrada por la curva descrita por sus coordenadas polares (ρ, θ), en donde $\rho = f(\theta)$, y las semirrectas con ángulos θ_1 y θ_2 ($\theta_1 < \theta_2$).

Proposición 3.11 *Si un objeto de masa m se mueve debido a la acción de un campo de fuerzas gravitatorias (determinadas por otro objeto de masa M), entonces el movimiento del objeto de masa m satisface la Segunda Ley de Kepler, es decir, el vector de posición del objeto "barre" áreas iguales en tiempos iguales.*

Demostración. Sean $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ la función que asigna la posición del objeto (de masa m) al tiempo t (y cuyo movimiento se realiza en el plano XY), $\theta(t)$ la función que nos da el ángulo del vector de posición $\gamma(t)$, y $f(\theta)$ la función que determina la coordenada ρ del vector de posición en términos del ángulo θ.

Si ahora definimos

$$\tilde{u}(\theta) = (\cos(\theta), \operatorname{sen}(\theta), 0),$$

tendremos que $\gamma(t)$ (en coordenadas cartesianas) estará dada por

$$\gamma(t) = f(\theta(t))\tilde{u}(\theta(t)),$$

de modo que

$$\gamma'(t) = f(\theta(t))\tilde{u}'(\theta(t))\theta'(t) + f'(\theta(t))\theta'(t)\tilde{u}(\theta(t)).$$

Si ahora sustituimos las dos últimas identidades en la expresión 3.10, tenemos que

$$\begin{aligned}
\hat{c} &= \gamma(t) \times \gamma'(t) \\
&= (f(\theta(t))\tilde{u}(\theta(t))) \times (f(\theta(t))\tilde{u}'(\theta(t))\theta'(t) + f'(\theta(t))\theta'(t)\tilde{u}(\theta(t))) \\
&= (f(\theta(t))\tilde{u}(\theta(t))) \times (f(\theta(t))\tilde{u}'(\theta(t))\theta'(t)) \\
&\quad + (f(\theta(t))\tilde{u}(\theta(t))) \times (f'(\theta(t))\theta'(t)\tilde{u}(\theta(t))) \\
&= f^2(\theta(t))\theta'(t)(\tilde{u}(\theta(t)) \times \tilde{u}'(\theta(t)))
\end{aligned}$$

y por tanto, dado que

$$\begin{aligned}
\tilde{u}(\theta(t)) \times \tilde{u}'(\theta(t))) &= (\cos(\theta(t)), \operatorname{sen}(\theta(t)), 0) \times (-\operatorname{sen}(\theta(t)), \cos(\theta(t)), 0) \\
&= (0, 0, 1)
\end{aligned}$$

para toda t, deducimos que

$$\begin{aligned}
\|\hat{c}\| &= \left\| f^2(\theta(t))\theta'(t)(\tilde{u}(\theta(t)) \times \tilde{u}'(\theta(t))) \right\| \\
&= f^2(\theta(t))\theta'(t) \left\| \tilde{u}(\theta(t)) \times \tilde{u}'(\theta(t)) \right\| \\
&= f^2(\theta(t))\theta'(t),
\end{aligned}$$

de donde obtenemos que

$$\theta'(t) = \frac{\|\hat{c}\|}{f^2(\theta(t))}.$$

Por otra parte, por la regla de la cadena (para funciones de $\mathbb{R}$ en $\mathbb{R}$) y el Teorema Fundamental del Cálculo, concluimos que

$$\begin{aligned}(A\circ\theta)'(t) &= A'(\theta(t))\theta'(t)\\ &= \left(\frac{1}{2}f^2(\theta(t))\right)\left(\frac{\|\hat{c}\|}{f^2(\theta(t))}\right)\\ &= \frac{\|\hat{c}\|}{2}.\end{aligned}$$

es decir, que $(A\circ\theta)'(t)$ es constante. ■

Si la fuerza $\hat{F}$ que se ejerce sobre el objeto en la posición $\gamma(t)$ no es de tipo gravitatorio, sino que es de la forma $\hat{F}(\gamma(t)) = \alpha(t)\gamma(t)$ (donde $\alpha(t)$ es un escalar que depende de t), las dos últimas proposiciones se siguen cumpliendo. Se dice que estos campos son de tipo "*central*". Dejamos como un problema al lector realizar las pruebas correspondientes.

3.6. Problemas

1. Sea $R\subset\mathbb{R}^2$ la recta cuya ecuación cartesiana es $ax+by+c=0$ (con $a^2+b^2>0$). Muestre que:

   a) si $\hat{x}_0=(x_0,y_0)$ y $\hat{x}_1=(x_1,y_1)$ son dos puntos diferentes que pertenecen a R, entonces la función $\gamma:\mathbb{R}\to\mathbb{R}^2$ dada por

   $$\gamma(t)=\hat{x}_0+t(\hat{x}_1-\hat{x}_0)$$

   es una parametrización de R.

   b) si $\hat{x}_0\in R$ y $\hat{u}=(u_1,u_2)\neq\hat{0}$ es un vector paralelo a la recta R (es decir, que $au_1+bu_2=0$), entonces la función $\gamma:\mathbb{R}\to\mathbb{R}^2$ dada por

   $$\gamma(t)=\hat{x}_0+t\hat{u}$$

   es una parametrización de R.

2. Sea $R \subset \mathbb{R}^3$ la recta determinada por la intersección de los planos $ax + by + cz + d = 0$ y $\tilde{a}x + \tilde{b}y + \tilde{c}z + \tilde{d} = 0$. Muestre que:

   a) si $\hat{x}_0 = (x_0, y_0, z_0)$ y $\hat{x}_1 = (x_1, y_1, z_1)$ son dos puntos diferentes que pertenecen a R, entonces la función $\gamma : \mathbb{R} \to \mathbb{R}^3$ dada por

   $$\gamma(t) = \hat{x}_0 + t(\hat{x}_1 - \hat{x}_0)$$

   es una parametrización de R.

   b) si $\hat{x}_0 \in R$ y $\hat{u} = (u_1, u_2, u_3) \neq \hat{0}$ es un vector paralelo a la recta R (es decir, que $au_1 + bu_2 + cu_3 = 0$ y $\tilde{a}u_1 + \tilde{b}u_2 + \tilde{c}u_3 = 0$), entonces la función $\gamma : \mathbb{R} \to \mathbb{R}^3$ dada por

   $$\gamma(t) = \hat{x}_0 + t\hat{u}$$

   es una parametrización de R.

3. Sean $\hat{x}_0, \hat{x}_1 \in \mathbb{R}^n$ dos puntos diferentes y $\hat{u} \neq \hat{0} \in \mathbb{R}^n$, con $n > 3$.

   a) ¿cómo se define, por medio de ecuaciones cartesianas, a la recta que pasa por los puntos $\hat{x}_0$ y $\hat{x}_1 1\mathbb{R}^n$?

   b) ¿cómo se define, por medio de ecuaciones cartesianas, a la recta en la dirección del vector $\hat{u}$ que pasa por el punto $\hat{x}_0$?

   c) ¿cómo definiría, sin usar ecuaciones cartesianas, a la recta que pasa por los puntos $\hat{x}_0$ y $\hat{x}_1$, y a la recta que pasa por el punto $\hat{x}_0$ que está en la dirección determinada por el vector $\hat{u}$? (los dos problemas anteriores le pueden dar una pista).

4. Sobre la parte exterior de una circunferencia fija de radio a rueda (sin resbalar) otra circunferencia de radio b. Encuentre una función de $\mathbb{R}$ en $\mathbb{R}^2$ que describa el movimiento de un punto que se encuentre en la circunferencia exterior.

5. Pruebe la proposición 3.2 usando los resultados que se indican en el texto, y después sin usar la proposición 3.1.

6. Pruebe la proposición 3.3 sin usar funciones coordenadas ni la proposición 3.1.

7. Muestre que la curva descrita por la función $\gamma(t) = (\operatorname{sen}(2t), 2\operatorname{sen}^2(t), 2\cos(t))$ $(t \in \mathbb{R})$ pertenece a una esfera con centro en el origen. Calcule su rapidez y muestre que la proyección en el plano XY de su velocidad tiene norma constante.

8. Sea $f : I \subset \mathbb{R} \to \mathbb{R}$ y $\gamma(t) = (t, f(t))$ parametrización de la gráfica de f (G_f). Pruebe que:

   a) si f es derivable en $t_0 \in I$, entonces la parametrización γ también es derivable en t_0 y además $\gamma'(t_0) = (1, f'(t_0))$

   b) la recta tangente a G_f en el punto $(t, f(t_0))$ es una parametrización de la misma recta (tangente) dada por la ecuación $y = f'(t_0)(x - t_0) + f(t_0)$.

9. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ y $t_0 \in I$. Pruebe que las siguientes afirmaciones son equivalentes:

   a) la función γ es derivable en t_0

   b) existe un vector $\hat{l} \in \mathbb{R}^n$ tal que

   $$\lim_{t \to t_0} \frac{\gamma(t) - (\gamma(t_0) + (t - t_0)\,\hat{l})}{t - t_0} = \hat{0}$$

   (interprete geométricamente este límite)

   c) existe una función lineal $L : \mathbb{R} \to \mathbb{R}^n$ tal que

   $$\lim_{t \to t_0} \frac{\gamma(t) - (\gamma(t_0) + L\,(t - t_0))}{t - t_0} = \hat{0}$$

   identifique geométricamente al conjunto

   $$\{\gamma(t_0) + L\,(t - t_0) \in \mathbb{R}^n \mid t \in \mathbb{R}\}\,.$$

Muestre que, de la equivalencia entre los incisos (a) y (b) se concluye que $\hat{l} = \gamma'(t_0)$, de la equivalencia entre los incisos (b) y (c) que $L(1) = \hat{l}$, y de la equivalencia entre los incisos (c) y (a) que $\gamma'(t_0) = L(1)$.

10. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ derivable. Pruebe que: $\|\gamma(t)\|$ es constante si y sólo si $\gamma(t) \cdot \gamma'(t) = 0$ para toda $t \in I$. Interprete geométricamente.

11. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ derivable y $r(t) = \|\gamma(t)\|$. Si $r(t_0)$ es un máximo o mínimo local de r, pruebe que $\gamma(t_0) \cdot \gamma'(t_0) = 0$. Interprete geométricamente.

12. Sea $\gamma = (f, g) : [a, b] \subset \mathbb{R} \to \mathbb{R}^2$ continua en el intervalo $[a, b]$ y derivable en el intervalo (a, b) tal que $\gamma'(t) \neq \hat{0}$ para toda $t \in (a, b)$. Pruebe que existen $\xi \in (a, b)$ y $\lambda \in \mathbb{R}$ tales que

$$\gamma(b) - \gamma(a) = \lambda\gamma'(\xi).$$

13. Sea $\gamma : [a, b] \subset \mathbb{R} \to \mathbb{R}^n$ continua en el intervalo $[a, b]$ y derivable en el intervalo (a, b). Pruebe que existe $\xi \in (a, b)$ tal que

$$\|\gamma(b) - \gamma(a)\|^2 = (b - a)\gamma'(\xi) \cdot (\gamma(b) - \gamma(a)).$$

14. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$.

    a) Pruebe que si existen $\hat{x}_0, \hat{u} \in \mathbb{R}^n$ tales que $\gamma(t) = \hat{x}_0 + t\hat{u}$ para toda $t \in I$, entonces $\|\gamma'(t)\|$ es constante.

    b) Muestre con un ejemplo que el recíproco de la afirmación del inciso anterior es falsa.

    c) Si existen $\hat{x}_0, \hat{u} \in \mathbb{R}^n$ tales que $\gamma(t) = \hat{x}_0 + t\hat{u}$ para toda $t \in I$, pruebe que $\gamma''(t) = \hat{0}$ para toda $t \in I$.

    d) Si $\gamma''(t) = \hat{0}$ para toda $t \in I$, pruebe que existen $\hat{x}_0, \hat{u} \in \mathbb{R}^n$ tales que $\gamma(t) = \hat{x}_0 + t\hat{u}$ para toda $t \in I$.

    e) De un ejemplo en el que la función γ parametrice una recta y para la cual se satisfaga que $\gamma''(t) \neq \hat{0}$ para toda $t \in I$.
    Interprete las afirmaciones de los incisos anteriores partiendo del hecho de que γ describe el movimiento de un objeto.

15. Dada $\gamma : [a,b] \subset \mathbb{R} \to \mathbb{R}^n$ continua, con $\gamma = (\gamma_1, \ldots, \gamma_n)$, defina

$$\int_a^b \gamma(t)dt = \left(\int_a^b \gamma_1(t)dt, \ldots, \int_a^b \gamma_n(t)dt \right).$$

Pruebe que:

a) si $\hat{c} = (c_1 \ldots, c_n)$ es un vector constante, entonces

$$\int_a^b \hat{c} \cdot \gamma(t)dt = \hat{c} \cdot \int_a^b \gamma(t)dt$$

b)

$$\left\| \int_a^b \gamma(t)dt \right\| \leq \int_a^b \|\gamma(t)\| \, dt$$

(*sugerencia*: argumente por qué se satisface que

$$\gamma(u) \cdot \left(\int_a^b \gamma(t)dt \right) \leq \|\gamma(u)\| \left\| \int_a^b \gamma(t)dt \right\|$$

para toda $u \in [a,b]$; integre con respecto de u y use la identidad del primer inciso)

c) si γ tiene derivada continua, entonces

$$\|\gamma(b) - \gamma(a)\| \leq \int_a^b \left\|\gamma'(t)\right\| dt = l(\gamma).$$

Interprete geométricamente.

16. Considere las funciones $\gamma(t) = (\cos(t), \operatorname{sen}(t))$, $t \in [0, \pi/2]$, y $\sigma(u) = \left(\frac{1-u^2}{1+u^2}, \frac{2u}{1+u^2} \right)$, $u \in [0,1]$. Pruebe que γ es una reparametrización de σ.

17. Sea $\gamma : [a,b] \subset \mathbb{R} \to \mathbb{R}^n$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^n$.

a) Muestre que $l(\gamma) = b - a$.

b) Muestre un ejemplo en el que *la longitud* de C no coincida con la longitud determinada por γ (es decir, $l(\gamma)$).

18. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^n$. Pruebe que si C está contenida en una recta, entonces $k(s) = 0$ para toda $s \in I$. ¿Esta afirmación contradice lo que se pide ejemplificar en el inciso (e) del problema 14? Justifique su respuesta.

19. Pruebe el recíproco del problema anterior. Es decir, si $C \subset \mathbb{R}^n$ es una curva suave y $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ es una parametrización por longitud de arco de C tal que $k(s) = 0$ para toda $s \in I$, entonces C está contenida en una recta.

20. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^n$. Si $T(s)$ es el vector tangente unitario y $\theta(s, h)$ representa el ángulo formado por los vectores $T(s)$ y $T(s+h)$, pruebe que

$$k(s) = \lim_{h \to 0} \left| \frac{\theta(s, h)}{h} \right|$$

(*sugerencia*: use la ley de los cosenos).

21. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^3$. Si $B(s)$ es el vector binormal unitario y $\theta(s, h)$ representa el ángulo formado por $B(s)$ y $B(s + h)$, pruebe que

$$\tau(s) = \pm \lim_{h \to 0} \left| \frac{\theta(s, h)}{h} \right|.$$

22. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^3$. Si $\gamma''(s) \neq \hat{0}$ para toda $s \in I$, pruebe que:

$$\tau(s) = \frac{[\gamma'(s) \times \gamma''(s)] \cdot \gamma'''(s)}{\|\gamma''(s)\|^2}.$$

23. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^3$. Si $\tau(s) = 0$ para toda $s \in I$, entonces C está contenida en un plano.

24. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^3$. Pruebe que, si la curvatura $k(s)$ es constante (distinta de cero) y $\tau(s) = 0$ para toda $s \in I$, entonces la curva descrita por γ es (o está contenida en) una circunferencia (*sugerencia*: pruebe que el centro de curvatura en $\gamma(s)$ es el mismo para toda $s \in I$ usando la expresión para $N'(s)$ probada en la proposición 3.8).

25. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^3$. Pruebe que, si para cada $s \in I$ la recta normal a C dada por $\gamma(s)+tN(s)$, $t \in \mathbb{R}$, pasa por un punto fijo $\hat{x}_0 \in \mathbb{R}^3$, entonces C está contenida en una circunferencia con centro en $\hat{x}_0$ (*sugerencia*: use el problema 24).

26. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^2$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^2$. Pruebe que si la curvatura $k(s)$ es constante (distinta de cero) en cada punto $\gamma(s)$, entonces la curva descrita por γ es (o está contenida en) una circunferencia (*sugerencia*: use el problema 24).

27. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ una parametrización por longitud de arco de una curva $C \subset \mathbb{R}^n$. Dado $t_0 \in \mathbb{R}$ fijo, definimos $\tilde{\gamma} : J \subset \mathbb{R} \to \mathbb{R}^n$, con $J = \{s \in \mathbb{R} \mid t_0 - s \in I\}$, como $\tilde{\gamma}(s) = \gamma(t_0 - s)$. Pruebe que:

    a) $\tilde{\gamma}$ es una parametrización por longitud de arco de C

    b) si $T(s), \tilde{T}(s), N(s), \tilde{N}(s), k(s)$ y $\tilde{k}(s)$ son los vectores tangente unitario, normal unitario y la curvatura, correspondientes a las parametrizaciones γ y $\tilde{\gamma}$ respectivamente, entonces
    i) $\tilde{T}(s) = -T(t_0 - s)$ *ii)* $\tilde{N}(s) = N(t_0 - s)$

    iii) $\tilde{k}(s) = k(t_0 - s)$
    para cada $s \in J$

    c) si $n = 3$ y $B(s), \tilde{B}(s), \tau(s)$ y $\tilde{\tau}(s)$ son los vectores binormal unitario y la torsión, correspondientes a las parametrizaciones γ y $\tilde{\gamma}$ respectivamente, entonces
    i) $\tilde{B}(s) = -B(t_0 - s)$ *ii)* $\tilde{\tau}(s) = \tau(t_0 - s)$

para cada $s \in J$

d) si $n = 3$, se satisfacen las fórmulas de Frenet-Serret escribiendo sus elementos en términos de $\tilde{\gamma}$.

28. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una curva suave $C \subset \mathbb{R}^3$. Definimos $\tilde{T}(t) = \gamma'(t)/\left\|\gamma'(t)\right\|$ y, bajo el supuesto de que $\tilde{T}'(t) \neq \hat{0}$ para toda $t \in I$, hacemos $\tilde{N}(t) = \tilde{T}'(t)/\left\|\tilde{T}'(t)\right\|$ y $\tilde{B}(t) = \tilde{T}(t) \times \tilde{N}(t)$. Si $\tilde{\gamma}$ es la parametrización por longitud de arco de C que se construye en la proposición 3.5, y $T(s)$, $N(s)$ y $B(s)$ son los vectores determinados por $\tilde{\gamma}$ de acuerdo con la definición 3.10, pruebe que: $\tilde{T}(t) = T(\alpha(t)) = (T \circ \alpha)(t)$; $\tilde{N}(t) = N(\alpha(t)) = (N \circ \alpha)(t)$ y $\tilde{B}(t) = B(\alpha(t)) = (B \circ \alpha)(t)$, en donde α es la función de longitud de arco definida en la proposición 3.5.

29. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una curva suave $C \subset \mathbb{R}^3$. Sean $\tilde{\gamma}$, α, $\tilde{T}(t)$, $\tilde{N}(t)$ y $\tilde{B}(t)$ como en el problema 28. Pruebe que:

a) si $\tilde{k}(t) = \left\|\tilde{T}'(t)\right\|$, entonces

$$k(\alpha(t)) = \tilde{k}(t)/\left\|\gamma'(t)\right\|$$

(donde $k(\alpha(t)) = \|\tilde{\gamma}''(\alpha(t))\|$ es la curvatura de C para $s = \alpha(t)$, de acuerdo con la definición 3.9)

b) $\tilde{B}'(t) \cdot \tilde{B}(t) = 0 = \tilde{B}'(t) \cdot \tilde{T}(t)$

c) $\tilde{B}'(t)$ es un múltiplo escalar de $\tilde{N}(t)$

d) si denotamos por $-\tilde{\tau}(t)$ al número (del inciso anterior) tal que $\tilde{B}'(t) = -\tilde{\tau}(t)\tilde{N}(t)$, pruebe que $\tau(\alpha(t)) = \tilde{\tau}(t)/\left\|\gamma'(t)\right\|$ (donde $\tau(\alpha(t))$ es la torsión de C para $s = \alpha(t)$, de acuerdo con la definición 3.11)

e) $\tilde{N}'(t) = -\tilde{k}(t)\tilde{T}(t) + \tilde{\tau}(t)\tilde{B}(t)$.

30. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^n$ una curva suave. Pruebe que la curvatura k en cada punto $\gamma(t)$ está dada por:

$$\frac{\left(\|\gamma'(t)\|^2\|\gamma''(t)\|^2-(\gamma'(t)\cdot\gamma''(t))^2\right)^{1/2}}{\|\gamma'(t)\|^3}.$$

31. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una curva suave tal que $\gamma''(t) \neq \hat{0}$ para toda $t \in I$. Pruebe que la curvatura k y la torsión τ en cada punto $\gamma(t)$ están dadas por:

$$i)\ k = \frac{\|\gamma'(t)\times\gamma''(t)\|}{\|\gamma'(t)\|^3} \qquad ii)\ \tau = \frac{[\gamma'(t)\times\gamma''(t)]\cdot\gamma'''(t)}{\|\gamma'(t)\times\gamma''(t)\|^2}.$$

32. Sea $\gamma : I \subset \mathbb{R} \to \mathbb{R}^3$ una curva suave. Pruebe que si la curvatura k es constante (distinta de cero) y la torsión τ es cero en cada punto $\gamma(t)$, entonces la curva descrita por γ es (o está contenida en) una circunferencia.

33. Sea $f : [a,b] \to \mathbb{R}$ una función dos veces derivable. Pruebe que la curvatura en el punto $(x, f(x))$ de la gráfica de f está dada por

$$k(x) = \frac{|f''(x)|}{\left[1+(f'(x))^2\right]^{3/2}}.$$

34. Sea $\gamma(t) = (a\cos(\omega t), a\,\mathrm{sen}(\omega t), bt)$, $t \in \mathbb{R}$.

    a) Calcule la parametrización por longitud de arco de esta curva.

    b) Calcule los vectores $T(s), N(s)$ y $B(s)$ en cada punto de esta curva.

    c) Calcule la curvatura, el radio de curvatura y la torsión en cada punto de esta curva.

35. Un objeto gira (en el sentido de las manecillas del reloj) sobre una circunferencia centrada en el origen y de radio R con rapidez constante v. Si el objeto se desprende de la circunferencia en el punto

$(-R/\sqrt{2}, R/\sqrt{2})$, ¿en qué punto y en cuánto tiempo intersecta al eje Y? ¿Cuál debería ser la rapidez del objeto sobre la circunferencia para que alcance el mismo punto en la mitad del tiempo?

36. Un ratón se mueve con rapidez constante v sobre una circunferencia de radio R y un gato, también con rapidez constante v, persigue al ratón (empezando desde el centro de la circunferencia) de tal forma que el ratón, el gato y el centro de la circunferencia siempre son coliniales. ¿Alcanza el gato al ratón? ¿en qué punto? ¿en qué tiempo?

37. La posición (en $\mathbb{R}^3$) de un objeto de masa m está dada por la función $\gamma(t)$. Suponga que la fuerza $\hat{F}$ que se ejerce sobre el objeto en la posición $\gamma(t)$ (la cual produce su movimiento) es tal que $\hat{F}(\gamma(t)) = \alpha(t)\gamma(t)$ (donde $\alpha(t)$ es un escalar que depende de t). Pruebe que:

    a) el objeto se mueve sobre un plano (*sugerencia*: considere el producto $\gamma(t) \times \gamma'(t)$ y use la segunda ley de Newton),

    b) suponiendo que la curva descrita es cerrada, pruebe que el movimiento del objeto satisface la Segunda Ley de Kepler (*sugerencia*: proceda como en la prueba de la proposición 3.11).

Capítulo 4

La derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}$

El contenido de este capítulo será sin duda el primero en el que aparecerán conceptos realmente novedosos para el lector. Y aun cuando los primeros conceptos de derivación que definiremos para este tipo de funciones son muy "parecidos" al concepto de derivada de funciones de $\mathbb{R}$ en $\mathbb{R}$, la definición general de la derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}$ abrirá un camino hacia una visión más general de este concepto.

Antes de iniciar nuestro recorrido por estos temas, vamos a revisar algunos aspectos relacionados con los conjuntos en donde se encuentran las variables de las funciones con las que vamos a trabajar y la relación entre las varias representaciones que podemos hacer de éstos por medio del conjunto $\mathbb{R}^n$.

4.1. Un interludio de Álgebra Lineal

Como vimos en el capítulo 1, los objetos matemáticos que sirven para representar a las variables (independientes o dependientes) de las funciones que nos ocupan pertenecen a conjuntos que suelen estar dotados de una estructura algebraica que hace de ellos un cierto tipo de espacio que se conocen

con el nombre de *espacios vectoriales de dimensión finita sobre los números reales*.

Como ya también vimos, una de las características más importantes de este tipo de espacios con el que vamos a trabajar, y que por ahora denotaremos en general con la letra V, es que siempre hay varias maneras de elegir una colección finita de elementos de V, digamos $\{\hat{v}_1, \ldots, \hat{v}_n\}$, con la propiedad de que para cualquier otro elemento $\hat{v} \in V$ existen $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, únicos, tales que

$$\hat{v} = \lambda_1 \hat{v}_1 + \cdots + \lambda_n \hat{v}_n.$$

Como mencionamos en su momento (y como seguramente el lector ya sabe), este tipo de colecciones reciben el nombre de *base de* V, y la propiedad que las define es la que permite identificar al conjunto V con el conjunto $\mathbb{R}^n$. Esta identificación de V con $\mathbb{R}^n$ tiene la ventaja de que hace corresponder adecuadamente la estructura algebraica de V con la estructura algebraica que definimos para $\mathbb{R}^n$, estructura que por cierto, ¡convierte a $\mathbb{R}^n$ en un espacio vectorial!

Además de lo anterior, en el capítulo 1 también vimos algunos ejemplos específicos de espacios vectoriales, como es el caso de las flechas (del plano o el espacio) que parten de un punto fijo. Estos espacios particulares están dotados de estructuras geométricas (tales como el concepto de distancia o de ángulo), las cuales pudimos "trasladar" a $\mathbb{R}^n$, y consecuentemente a cualquier otro espacio vectorial que se pueda identificar con éste.

Observemos que si $\{\hat{v}_1, \ldots, \hat{v}_n\}$ es una base del espacio vectorial V, la $n-$ada con la que se identifica al vector $\hat{v}_i$ es aquella que tiene casi todas sus coordenadas 0, salvo por la $i-$ésima, que deberá ser 1; a esta $n-$ada la denotaremos por $\hat{e}_i$. Es decir que

$$\hat{e}_i := (0, \ldots, 0, 1, 0, \ldots, 0),$$

en donde el 1 está en la coordenada i.

Como seguramente el lector ya sabe (y si no, lo podrá probar muy fácilmente), la colección $\{\hat{e}_1, \ldots, \hat{e}_n\}$ es una base para $\mathbb{R}^n$ y se le conoce con el nombre de *base canónica*.

De acuerdo con los conceptos de magnitud (norma) y de ángulo (producto punto) que definimos en el capítulo 1, todos los elementos de esta base tienen magnitud 1 y son mutuamente perpendiculares. Es decir, $\|\hat{e}_i\| = 1$ para cada $i \in \{1, \ldots, n\}$ y $\hat{e}_i \cdot \hat{e}_j = 0$ para cada $i, j \in \{1, \ldots, n\}$, con $i \neq j$, o lo que es lo mismo,

$$\hat{e}_i \cdot \hat{e}_j = \begin{cases} 1 & \text{si } i = j \\ \\ 0 & \text{si } i \neq j \end{cases}$$

para cada $i, j \in \{1, \ldots, n\}$. En general, cuando los elementos de una base de un espacio vectorial se indentifican con $n-$adas que cumplen con estas dos características, decimos que esta base es *ortonormal*.

Como ya habíamos mencionado en el capítulo 1, este tipo de bases son las que se utilizan para construir (en el caso del plano o del espacio) un sistema de referencia cartesiano. Es por esta razón que en estos casos, cuando se dibujen los vectores (o flechas) que representen a una base de este tipo, habrá que hacerlo con estas características (mutuamente perpendiculares, y de la misma longitud). No sin cierto abuso de notación, usaremos los mismos $\hat{e}_1, \ldots, \hat{e}_n$ para nombrar a los vectores (o flechas) que representen geométricamente a la base canónica de $\mathbb{R}^n$.

Una vez establecidos los conceptos y notaciones anteriores, la situación principal que deseamos abordar en esta sección es la siguiente: supongamos que la variable independiente de una cierta función pertenece a un espacio vectorial V, y que las colecciones $\{\hat{v}_1, \ldots, \hat{v}_n\} \subset V$ y $\{\hat{v}'_1, \ldots, \hat{v}'_n\} \subset V$, ambas, son bases de V. Si identificamos al espacio V con $\mathbb{R}^n$ a través de la base $\{\hat{v}_1, \ldots, \hat{v}_n\}$ (en cuyo caso cada vector $\hat{v}_i$ se corresponderá con la n-ada $\hat{e}_i$), entonces, usando esta misma manera de identificar a V con $\mathbb{R}^n$, cada vector $\hat{v}'_i$ se corresponderá con alguna n-ada, que denotaremos por $\hat{e}'_i$, y se tendrá que la colección de n-adas $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ será una base (o sistema coordenado) de $\mathbb{R}^n$ (afirmación que seguramente el lector ya habrá probado en su curso de Álgebra Lineal).

Algo que es importante destacar es que, aun cuando la colección $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ sea una base de $\mathbb{R}^n$, ésta no tiene que ser necesariamente ortonormal. *Para los fines de este texto, supondremos que las colecciones* $\{\hat{v}_1, \ldots, \hat{v}_n\} \subset$

V y $\{\hat{v}'_1, \ldots, \hat{v}'_n\} \subset V$ se elegirán de tal manera que la colección $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ siempre sea una base ortonormal (o sistema coordenado cartesiano) de $\mathbb{R}^n$.

De esta forma, dadas las colecciones $\{\hat{v}_1, \ldots, \hat{v}_n\} \subset V$ y $\{\hat{v}'_1, \ldots, \hat{v}'_n\} \subset V$, la situación que tendremos (como ya habíamos mencionado desde el capítulo 1) es que habrá dos formas diferentes de identificar al espacio vectorial V con el conjunto (o espacio vectorial) $\mathbb{R}^n$. Esta doble identificación se traducirá en lo siguiente: así como para cada $\hat{v} \in V$ existirán dos formas de expresar a $\hat{v}$ como combinación lineal de las bases $\{\hat{v}_1, \ldots, \hat{v}_n\}$ y $\{\hat{v}'_1, \ldots, \hat{v}'_n\}$, para cada $\hat{x} \in \mathbb{R}^n$ (pensando a $\mathbb{R}^n$ más como un "representante" del espacio vectorial V, que como un conjunto de n-adas), consideraremos dos sistemas coordenados cartesianos en los cuales expresaremos a $\hat{x}$. Es decir, un vector $\hat{x} \in \mathbb{R}^n$ tendrá coordenadas $(x_1, \ldots, x_n)$ en el sistema coordenado cartesiano determinado por la base canónica $\{\hat{e}_1, \ldots, \hat{e}_n\}$, y coordenadas $(x'_1, \ldots, x'_n)$ en el otro sistema coordenado cartesiano determinado por la base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$. La figura 4.1 ilustra esta situación para el caso particular del espacio vectorial V formado por las flechas que parten de un punto fijo O, razón por la cual las bases $\{\hat{e}_1, \hat{e}_2\}$ y $\{\hat{e}'_1, \hat{e}'_2\}$ se "fijan" en el mismo punto.

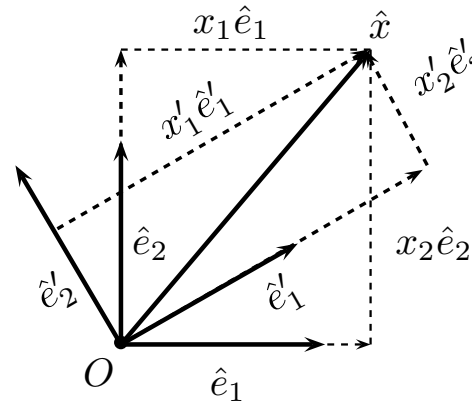


Figura 4.1: El vector $\hat{x}$ y sus correspondientes coordenadas (x_1, x_2) y (x'_1, x'_2) en las bases ortonormales $\{\hat{e}_1, \hat{e}_2\}$ y $\{\hat{e}'_1, \hat{e}'_2\}$, respectivamente.

En virtud de lo anterior, tendremos que una misma función f definida para los elementos de un subconjunto A del espacio vectorial V (y que en un abuso de lenguaje diremos que está definida para los elementos de un subconjunto A del espacio vectorial $\mathbb{R}^n$), ésta se podrá poner en términos de unas ciertas variables (o coordenadas) $x_1, \ldots, x_n$, o de otras variables (o coordenadas) $x'_1, \ldots, x'_n$.

Aun cuando los conceptos de derivabilidad con los que vamos a trabajar en este capítulo serán definidos sin tener que recurrir a las coordenadas determinadas por alguna base, para efectos de la realización de cálculos específicos sí será necesario recurrir a alguna de éstas. Lo que nos proponemos en este capítulo es dejar clara la relación que existirá entre los cálculos realizados con unas cooredenadas $x_1, \ldots, x_n$ o con otras coordenadas $x'_1, \ldots, x'_n$. Para alcanzar este objetivo, empezaremos por establecer la relación que existe entre dos conjuntos de coordenadas (cartesianas) de un mismo elemento $\hat{x} \in \mathbb{R}^n$.

Si, como dijimos antes, $(x_1, \ldots, x_n)$ son las coordenadas de $\hat{x}$ en el sistema coordenado cartesiano determinado por la base canónica $\{\hat{e}_1, \ldots, \hat{e}_n\}$, y $(x'_1, \ldots, x'_n)$ son otras coordenadas en un sistema coordenado cartesiano determinado por una base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$, nuestro objetivo es mostrar cómo se obtienen unas coordenadas a partir de las otras; es decir, si conocemos las coordenadas $(x_1, \ldots, x_n)$, cómo podemos obtener las coordenadas $(x'_1, \ldots, x'_n)$ y recíprocamente, si conocemos las coordenadas $(x'_1, \ldots, x'_n)$, cómo podemos obtener las coordenadas $(x_1, \ldots, x_n)$.

Como seguramente el lector ya sabrá, cada uno de estos problemas es un problema de *cambio de coordenadas*. También sabrá que se resuelve de la siguiente manera: supongamos primero que tenemos las coordenadas $(x'_1, \ldots, x'_n)$ de $\hat{x} \in \mathbb{R}^n$ en la base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ y que deseamos conocer sus coordenadas $(x_1, \ldots, x_n)$ en la base canónica. Para ello, bastará con saber cuáles son las coordenadas de cada vector $\hat{e}'_i$ en la base canónica $\{\hat{e}_1, \ldots, \hat{e}_n\}$ (información con la que se cuenta la mayoría de las veces).

Si suponemos que

$$\begin{aligned}\hat{e}'_i &= \left(a_1^{(i)}, \ldots, a_n^{(i)}\right)\\ &= a_1^{(i)}\hat{e}_1 + \cdots + a_n^{(i)}\hat{e}_n\end{aligned}$$

para cada $i \in \{1, \ldots, n\}$, entonces

$$\begin{aligned}\hat{x} &= x'_1\hat{e}'_1 + \cdots + x'_n\hat{e}'_n\\ &= x'_1\left(a_1^{(1)}\hat{e}_1 + \cdots + a_n^{(1)}\hat{e}_n\right) +\\ &\vdots\\ &+ x'_n\left(a_1^{(n)}\hat{e}_1 + \cdots + a_n^{(n)}\hat{e}_n\right)\\ &= \left(a_1^{(1)}, \ldots, a_1^{(n)}\right)\cdot(x'_1, \ldots, x'_n)\hat{e}_1 +\\ &\vdots\\ &+ \left(a_n^{(1)}, \ldots, a_n^{(n)}\right)\cdot(x'_1, \ldots, x'_n)\hat{e}_n,\end{aligned}$$

de donde concluimos que la i-ésima coordenada del vector $\hat{x}$ en la base canónica estará dada por

$$x_i = \left(a_i^{(1)}, \ldots, a_i^{(n)}\right)\cdot(x'_1, \ldots, x'_n)$$

para cada $i \in \{1, \ldots, n\}$, o equivalentemente y escrito usando matrices, que

$$\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} a_1^{(1)} & \cdots & a_n^{(1)} \\ \vdots & \ddots & \vdots \\ a_1^{(n)} & \cdots & a_n^{(n)} \end{bmatrix}. \tag{4.1}$$

Como el lector se podrá imaginar, para resolver el problema recíproco (obtener las coordenadas $(x'_1,\ldots,x'_n)$ a partir de las coordenadas $(x_1, \ldots, x_n)$), será suficiente con calcular las coordenadas de cada vector $\hat{e}_i$ en el sistema cartesiano determinado por la base $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$.

Es decir, si ahora

$$\begin{aligned}\hat{e}_i &= \left(b_1^{(i)}, \ldots, b_n^{(i)}\right) \\ &= b_1^{(i)}\hat{e}'_1 + \cdots + b_n^{(i)}\hat{e}'_n\end{aligned}$$

para cada $i \in \{1, \ldots, n\}$, entonces

$$\begin{aligned}\hat{x} &= x_1\hat{e}_1 + \cdots + x_n\hat{e}_n \\ &= x_1\left(b_1^{(1)}\hat{e}'_1 + \cdots + b_n^{(1)}\hat{e}'_n\right) + \\ &\vdots \\ &+ x_n\left(b_1^{(n)}\hat{e}'_1 + \cdots + b_n^{(n)}\hat{e}'_n\right) \\ &= \left(b_1^{(1)}, \ldots, b_1^{(n)}\right)\cdot(x_1, \ldots, x_n)\hat{e}'_1 + \\ &\vdots \\ &+ \left(b_n^{(1)}, \ldots, b_n^{(n)}\right)\cdot(x_1, \ldots, x_n)\hat{e}'_n,\end{aligned}$$

de donde concluimos que la $i-$ésima coordenada del vector $\hat{x}$ en la base $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ estará dada por

$$x'_i = \left(b_i^{(1)}, \ldots, b_i^{(n)}\right)\cdot(x_1, \ldots, x_n)$$

para cada $i \in \{1, \ldots, n\}$, o lo que es lo mismo, escrito en forma matricial, que

$$\begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} b_1^{(1)} & \cdots & b_n^{(1)} \\ \vdots & \ddots & \vdots \\ b_1^{(n)} & \cdots & b_n^{(n)} \end{bmatrix}. \tag{4.2}$$

Si llamamos

$$M = \begin{bmatrix} a_1^{(1)} & \cdots & a_n^{(1)} \\ \vdots & \ddots & \vdots \\ a_1^{(n)} & \cdots & a_n^{(n)} \end{bmatrix}$$

y

$$M' = \begin{bmatrix} b_1^{(1)} & \cdots & b_n^{(1)} \\ \vdots & \ddots & \vdots \\ b_1^{(n)} & \cdots & b_n^{(n)} \end{bmatrix}$$

se prueba que:

1. cada matriz M y M' es la inversa una de la otra ($M' = M^{-1}$),
2. la inversa de cada una de ellas es su propia transpuesta ($M^{-1} = M^t$), y
3. ambas matrices tienen determinante ± 1.

Nótese que de las propiedades 1 y 2 se deduce que

$$b_j^{(i)} = a_i^{(j)} \tag{4.3}$$

para todas $i, j \in \{1, \ldots, n\}$.

Las matrices que tienen las características anteriores son conocidas con el nombre de *matrices ortonormales* (por razones que seguramente quedan claras). Si el lector aún no conoce la prueba de estas afirmaciones, muy probablemente pronto las verá en su curso de Álgebra Lineal).

Resumiendo la discusión anterior, con base en las identidades 4.1, 4.2 y 4.3, concluimos que: si cada vector $\hat{e}'_i$ tiene coordenadas $\left(a_1^{(i)}, \ldots, a_n^{(i)}\right)$ ($i \in \{1, \ldots, n\}$) en el sistema determinado por la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$, y un vector $\hat{x}$ tiene coordenadas $(x_1, \ldots, x_n)$ en el mismo sistema, entonces las coordenadas $(x'_1, \ldots, x'_n)$ de $\hat{x}$ (en el sistema determinado por la base $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$) están dadas por la identidad (escrita en forma matricial)

$$\begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_1^{(1)} & \cdots & a_1^{(n)} \\ \vdots & \ddots & \vdots \\ a_n^{(1)} & \cdots & a_n^{(n)} \end{bmatrix}. \tag{4.4}$$

Análogamente, si cada vector $\hat{e}_i$ tiene coordenadas $\left(b_1^{(i)}, \ldots, b_n^{(i)}\right)$ ($i \in \{1, \ldots, n\}$) en el sistema determinado por la base $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$, y un vector

$\hat{x}$ tiene coordenadas $(x'_1, \ldots, x'_n)$ en el mismo sistema, entonces las coordenadas $(x_1, \ldots, x_n)$ de $\hat{x}$ (en el sistema determinado por la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$) están dadas por la identidad (escrita en forma matricial)

$$\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} b_1^{(1)} & \cdots & b_1^{(n)} \\ \vdots & \ddots & \vdots \\ b_n^{(1)} & \cdots & b_n^{(n)} \end{bmatrix}. \tag{4.5}$$

Nótese que, de acuerdo con las identidades 4.3, también se tiene que

$$\begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} b_1^{(1)} & \cdots & b_n^{(1)} \\ \vdots & \ddots & \vdots \\ b_1^{(n)} & \cdots & b_n^{(n)} \end{bmatrix} \tag{4.6}$$

y

$$\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} a_1^{(1)} & \cdots & a_n^{(1)} \\ \vdots & \ddots & \vdots \\ a_1^{(n)} & \cdots & a_n^{(n)} \end{bmatrix}, \tag{4.7}$$

identidades que también nos serán muy útiles.

Una consecuencia muy importante de las identidades anteriores tiene que ver con los conceptos de producto punto y norma euclideana, los cuales se definieron en el capítulo 1 en términos de coordenadas. Como se recordará, si $(x_1, \ldots, x_n)$ y $(y_1, \ldots, y_n)$ son las coordenadas de $\hat{x}$ y $\hat{y}$ respectivamente, en el sistema cartesiano determinado por la base canónica $\{\hat{e}_1, \ldots, \hat{e}_n\}$, definimos que

$$\hat{x} \cdot \hat{y} = x_1 y_1 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i.$$

Si ahora suponemos que $(x'_1, \ldots, x'_n)$ y $(y'_1, \ldots, y'_n)$ son las coordenadas de $\hat{x}$ y $\hat{y}$ respectivamente, en un sistema coordenado cartesiano determinado por otra base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$, de acuerdo con la identidad 4.4 (y

denotando por Id_n a la matriz identidad de $n \times n$) se tiene que

$$\begin{aligned}
\left[x_1'y_1' + \cdots + x_n'y_n'\right] &= \begin{bmatrix} x_1' & \cdots & x_n' \end{bmatrix} \begin{bmatrix} y_1' \\ \vdots \\ y_n' \end{bmatrix} \\
&= \begin{bmatrix} x_1' & \cdots & x_n' \end{bmatrix} \begin{bmatrix} y_1' & \cdots & y_n' \end{bmatrix}^t \\
&= \left(\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} M^t\right) \left(\begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix} M^t\right)^t \\
&= \left(\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} M^t\right) \left(\left(M^t\right)^t \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^t\right) \\
&= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \left(M^t M\right) \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^t \\
&= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \left(M^{-1} M\right) \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^t \\
&= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} (Id_n) \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^t \\
&= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^t \\
&= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \\
&= \left[x_1y_1 + \cdots + x_ny_n\right].
\end{aligned}$$

Por lo tanto, se tiene que

$$x_1'y_1' + \cdots + x_n'y_n' = x_1y_1 + \cdots + x_ny_n.$$

De esta última identidad concluimos que el valor de $\hat{x} \cdot \hat{y}$ no depende del sistema coordenado cartesiano que se esté usando para identificar a los vectores $\hat{x}$ y $\hat{y}$, y si recordamos que

$$\|\hat{x}\| = \sqrt{\hat{x} \cdot \hat{x}},$$

también concluimos que el valor de la norma de un vector $\hat{x} \in \mathbb{R}^n$ no depende de las coordenadas (cartesianas) que se usen. Por ahora nos será suficiente con saber lo anterior y conocer las relaciones establecidas en 4.4, 4.5, 4.6 y 4.7.

Para concluir esta sección, daremos un

Ejemplo 4.1 *En el plano, una vez establecido un origen O, considere el sistema coordenado cartesiano determinado por dos vectores $\{\hat{e}_1, \hat{e}_2\}$ (y que por lo mismo tendremos que dibujar mutuamente perpendiculares y de la misma longitud), y el sistema cartesiano determinado por los vectores $\{\hat{e}_\theta, \hat{e}_\theta^\perp\}$, en donde $\hat{e}_\theta$ se obtiene de girar (en el sentido contrario a las manecillas del reloj) θ radianes al vector $\hat{e}_1$, y $\hat{e}_\theta^\perp$ se obtiene de girar (en el sentido contrario a las manecillas del reloj) $\pi/2$ radianes al vector $\hat{e}_\theta$ (ver figura 4.2).*

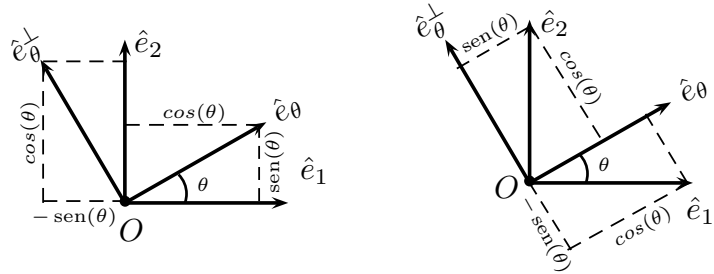


Figura 4.2: Las bases ortonormales $\{\hat{e}_1, \hat{e}_2\}$ y $\{\hat{e}_\theta, \hat{e}_\theta^\perp\}$ del ejemplo 4.1.

Por la manera en que construimos los vectores $\hat{e}_\theta$ y $\hat{e}_\theta^\perp$, sabemos que éstos forman una base ortonormal y que sus coordenadas en el sistema determinado por $\{\hat{e}_1, \hat{e}_2\}$ son $(\cos(\theta), \mathrm{sen}(\theta))$ para $\hat{e}_\theta$, y $(-\,\mathrm{sen}(\theta), \cos(\theta))$ para $\hat{e}_\theta^\perp$. Es decir, se tiene que

$$\begin{aligned} \hat{e}_\theta &= \cos(\theta)\hat{e}_1 + \mathrm{sen}(\theta)\hat{e}_2 \\ \hat{e}_\theta^\perp &= -\,\mathrm{sen}(\theta)\hat{e}_1 + \cos(\theta)\hat{e}_2. \end{aligned} \tag{4.8}$$

De esta forma, si $\hat{x}$ es un vector del plano que parte del punto O y que tiene coordenadas (x', y') en el sistema coordenado determinado por $\{\hat{e}_\theta, \hat{e}_\theta^\perp\}$, de acuerdo con la identidad 4.1, las coordenadas (x, y) de $\hat{x}$ en el sistema coordenado determinado por la base $\{\hat{e}_1, \hat{e}_2\}$ estarán dadas por la identidad de matrices

$$\begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} \cos(\theta) & \mathrm{sen}(\theta) \\ -\,\mathrm{sen}(\theta) & \cos(\theta) \end{bmatrix}.$$

Es decir

$$x = \cos(\theta)x' - \operatorname{sen}(\theta)y'$$
$$y = \operatorname{sen}(\theta)x' + \cos(\theta)y'.$$

Recíprocamente, si ahora observamos que

$$\hat{e}_1 = \cos(\theta)\hat{e}_\theta - \operatorname{sen}(\theta)\hat{e}_\theta^\perp$$
$$\hat{e}_2 = \operatorname{sen}(\theta)\hat{e}_\theta + \cos(\theta)\hat{e}_\theta^\perp$$

(identidades que podemos obtener "despejando" los vectores $\hat{e}_1$ y $\hat{e}_2$ de las identidades 4.8), de acuerdo con la identidad 4.2, si (x, y) son las coordenadas de $\hat{x}$ en el sistema coordenado determinado por la base $\{\hat{e}_1, \hat{e}_2\}$, las coordenadas (x', y') en el sistema coordenado determinado por $\{\hat{e}_\theta, \hat{e}_\theta^\perp\}$ estarán dadas por la identidad de matrices

$$\begin{bmatrix} x' & y' \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\operatorname{sen}(\theta) \\ \operatorname{sen}(\theta) & \cos(\theta) \end{bmatrix}$$

o lo que es lo mismo, por

$$x' = \cos(\theta)x + \operatorname{sen}(\theta)y$$
$$y' = -\operatorname{sen}(\theta)x + \cos(\theta)y.$$

Nótese que en este ejemplo se confirma que las matrices

$$M = \begin{bmatrix} \cos(\theta) & -\operatorname{sen}(\theta) \\ \operatorname{sen}(\theta) & \cos(\theta) \end{bmatrix} \quad y \quad M' = \begin{bmatrix} \cos(\theta) & \operatorname{sen}(\theta) \\ -\operatorname{sen}(\theta) & \cos(\theta) \end{bmatrix}$$

tienen las propiedades que habíamos mencionado.

4.2. La derivada direccional

Antes de entrar de lleno en la definición del concepto de derivada direccional, es importante mencionar que de aquí en adelante (salvo que se diga lo

contrario) supondremos que las funciones con las que trabajeremos están definidas sobre un conjunto abierto y conexo $U \subset \mathbb{R}^n$.

Sea pues $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Recordemos que el concepto de derivación de una función f de $\mathbb{R}$ en $\mathbb{R}$ en un punto x_0 de su dominio consiste básicamente en los siguientes pasos: calcular "el cambio" de los valores de f en x y en x_0, es decir, calcular $f(x) - f(x_0)$, en donde $x \neq x_0$; calcular "el cambio" que hay entre x y x_0 es decir, calcular $x - x_0$; calcular la "razón" (o "proporción") entre estas dos cantidades, es decir

$$\frac{f(x) - f(x_0)}{x - x_0};$$

y finalmente determinar si estas "razones" (o "proporciones") tienen un valor límite cuando la variable x se "aproxima" (o "tiende") al valor x_0. Como seguramente el lector ya habrá notado, no podemos copiar este procedimiento al caso de las funciones que nos ocupan. En efecto, si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0, \hat{x} \in U$, $\hat{x} \neq \hat{x}_0$, el "cambio" de los valores de la función en $\hat{x}$ y en $\hat{x}_0$ (es decir, $f(\hat{x}) - f(\hat{x}_0)$) y el "cambio" entre $\hat{x}$ y $\hat{x}_0$, es decir, $\hat{x} - \hat{x}_0$, son de naturaleza distinta; mientras que la primera de estas cantidades es un número real ($f(\hat{x}) - f(\hat{x}_0)$), la otra ($\hat{x} - \hat{x}_0$) es un vector, de tal forma que no tenemos una manera de calcular la "razón (o proporción)" entre dichas cantidades. Sin embargo, no todo está perdido y podemos adecuar la idea original, de alguna manera, al tipo de funciones que nos ocupa.

Dado que estamos suponiendo que U es un conjunto abierto, si $\hat{x}_0 \in U$ y $\hat{u}$ es un vector fijo que no sea el vector $\hat{0}$, los vectores de la forma $\hat{x} = \hat{x}_0 + h\hat{u}$, con $h \in \mathbb{R}$, tienen la particularidad de seguir perteneciendo a U si h es "suficientemente pequeña", aunque cabe aclarar que esta "pequeñez" de h también dependerá de la magnitud del vector $\hat{u}$. Más específicamente, sabemos que existe $r > 0$ tal que si $\|\hat{x} - \hat{x}_0\| < r$, entonces $\hat{x} \in U$, de modo que si tomamos $\hat{x} = \hat{x}_0 + h\hat{u}$, este punto pertenecerá a U si

$$\begin{aligned} r &> \|\hat{x} - \hat{x}_0\| \\ &= \|\hat{x}_0 + h\hat{u} - \hat{x}_0\| \\ &= \|h\hat{u}\| \\ &= |h| \, \|\hat{u}\| , \end{aligned}$$

es decir, si $|h| < r/\|\hat{u}\|$.

Por otra parte, si observamos que

$$\|\hat{x} - \hat{x}_0\| = |h|\,\|\hat{u}\|$$

vemos que la magnitud del "cambio" entre $\hat{x}$ y $\hat{x}_0$ está determinada tanto por h, como por la magnitud del vector $\hat{u}$.

Por las razones que acabamos de exponer, si al tomar $\hat{x} = \hat{x}_0 + h\hat{u}$ pedimos que el vector $\hat{u}$ tenga magnitud 1 lograremos las siguientes ventajas: primera, que la pertenencia del vector $\hat{x}$ al conjunto U sólo dependerá del número h (bastará con que $|h| < r$ para que $\hat{x} = \hat{x}_0 + h\hat{u}$ pertenezca a la bola $B_r(\hat{x}_0)$, la que a su vez está contenida en U); y segunda, el "cambio" entre $\hat{x}$ y $\hat{x}_0$ (es decir el vector $\hat{x} - \hat{x}_0$) estará completamente determinado por el número h (¡no sólo su magnitud! ¡también su dirección!).

Con base en los supuestos anteriores, el cociente

$$\frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h} \tag{4.9}$$

será una medida de la "razón (o proporción)" entre el cambio de los valores de la función (en los puntos $\hat{x}_0 + h\hat{u}$ y $\hat{x}_0$) y el cambio entre dichos valores de la variable, el cual está determinado por el número h.

Aquí es importante enfatizar que los puntos sobre los cuales estamos evaluando a f (para compararlos con su valor en $\hat{x}_0$) están restringidos a aquellos que son de la forma $\hat{x}_0 + h\hat{u}$, es decir, sólo estamos evaluando a f sobre los puntos que están en la recta que pasa por $\hat{x}_0$ y cuya dirección está determinada por el vector $\hat{u}$ (¡y que además se quedan contenidos en U!).

Por esta razón, si existe el límite del cociente 4.9 cuando h tiende a 0, diremos que la función f es derivable en el punto $\hat{x}_0$ en la dirección del vector $\hat{u}$, y al valor de este límite le llamaremos *la derivada direccional de f en $\hat{x}_0$ en la dirección del vector $\hat{u}$*. Dejamos plasmada la discusión anterior en la siguiente

Definición 4.1 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Decimos que f es derivable en el punto $\hat{x}_0$ en la dirección del vector $\hat{u}$, si*

$$\lim_{h\to 0} \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h}$$

existe, en cuyo caso llamaremos al valor de dicho límite la derivada direccional de f en $\hat{x}_0$ en la dirección de $\hat{u}$, y la denotaremos por $D_{\hat{u}}f(\hat{x}_0)$, es decir

$$D_{\hat{u}}f(\hat{x}_0) := \lim_{h\to 0} \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h}$$

Como dijimos anteriormente, en términos generales este concepto de derivada es una medida de la razón de cambio entre el valor de la función f en $\hat{x}_0$ y el valor en puntos de la recta que pasa por $\hat{x}_0$ y cuya dirección está determinada por el vector $\hat{u}$. En algunos casos, a semejanza de lo que sucede con las funciones de $\mathbb{R}$ en $\mathbb{R}$, dicha derivada también tiene una interpretación geométrica importante.

Para el caso particular de $\mathbb{R}^2$, si nos fijamos en la curva que obtenemos al intersecar la gráfica de f con el plano perpendicular al plano XY que contenga a la recta que mencionamos anteriormente (ver figura 4.3), la derivada direccional de f en $\hat{x}_0$ en la dirección de $\hat{u}$ puede interpretarse como la pendiente de la recta tangente a dicha curva en el punto $(\hat{x}_0, f(\hat{x}_0))$.

Sin duda lo siguiente que tendremos que hacer será dar un ejemplo que ilustre este concepto, pero antes de ello convendrá hacer algunas observaciones importantes. Como el lector podrá notar, en la definición que acabamos de dar no fue necesario recurrir a ningún sistema coordenado específico.

Por otra parte, como para hacer cálculos específicos sí será necesario expresar a f en términos de algunas coordenadas, en el siguiente ejemplo no sólo vamos a mostrar cómo se calcula la derivada direccional de f en algún punto $\hat{x}_0$ y en alguna dirección $\hat{u}$, sino que mostraremos que dichos cálculos son idependientes (cuando menos para este ejemplo específico) del sistema coordenado que usemos.

Ejemplo 4.2 *Sean $\hat{x}_0, \hat{u} \in \mathbb{R}^2$ tales que en la base canónica tienen coordenadas (x_0, y_0) y (u_1, u_2), respectivamente. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ tal que, si un*

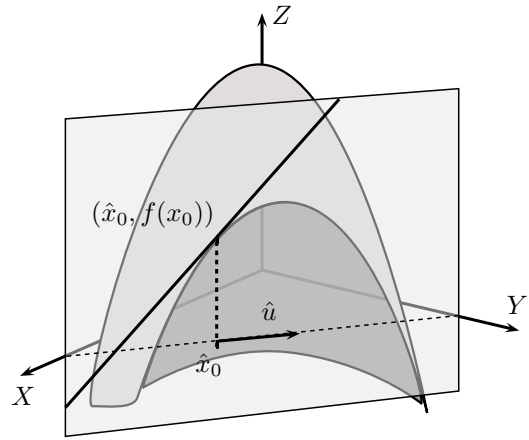


Figura 4.3: En $\mathbb{R}^2$, si intersecamos la gráfica de f con el plano perpendicular al plano XY que contenga a la recta que pasa por $\hat{x}_0$ en la dirección de $\hat{u}$, obtenemos una curva. Podemos interpretar la derivada $D_{\hat{u}}f(\hat{x}_0)$ como la pendiente de la recta tangente a dicha curva en el punto $(\hat{x}_0, f(\hat{x}_0))$.

punto $\hat{x} \in \mathbb{R}^2$ *tiene coordenadas* (x, y) *en la misma base, entonces el valor de* f *en* $\hat{x}$ *está dado por*

$$\begin{aligned} f(\hat{x}) &= f(x, y) \\ &= x^2 + y^2. \end{aligned}$$

Primero vamos a calcular $D_{\hat{u}}f(\hat{x}_0)$*. De acuerdo con la definición 4.1, tenemos que*

$$\begin{aligned} D_{\hat{u}}f(\hat{x}_0) &= \lim_{h\to 0} \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h} \\ &= \lim_{h\to 0} \frac{f(x_0 + hu_1, y_0 + hu_2) - f(x_0, y_0)}{h} \\ &= \lim_{h\to 0} \frac{(x_0 + hu_1)^2 + (y_0 + hu_2)^2 - (x_0^2 + y_0^2)}{h} \end{aligned}$$

$$\begin{aligned}
&= \lim_{h\to 0} \frac{2x_0hu_1 + (hu_1)^2 + 2y_0hu_2 + (hu_2)^2}{h} \\
&= \lim_{h\to 0} \left(2x_0u_1 + 2y_0u_2 + hu_1^2 + hu_2^2\right) \\
&= 2(x_0u_1 + y_0u_2).
\end{aligned}$$

Supongamos ahora que tenemos otro sistema coordenado determinado por una base $\{\hat{e}'_1, \hat{e}'_2\}$ *tal que el cambio de coordenadas de este sistema al sistema canónico está dado por la matriz*

$$M = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Es decir, si un punto $\hat{x}$ *tiene coordenadas* (x', y') *en la base* $\{\hat{e}'_1, \hat{e}'_2\}$*, entonces sus coordenadas* (x, y) *en la base canónica estarán dadas por*

$$\begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} x' & y' \end{bmatrix} M = \begin{bmatrix} x' & y' \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

o lo que es lo mismo, que

$$\begin{aligned}
x &= a_{11}x' + a_{21}y' \\
y &= a_{12}x' + a_{22}y'.
\end{aligned}$$

De esta forma, el valor de f *en* $\hat{x}$ *expresado en términos de sus coordenadas* (x', y')*, ahora estará dado por*

$$\begin{aligned}
f(x', y') &= f(\hat{x}) \\
&= f(x, y) \\
&= x^2 + y^2 \\
&= (a_{11}x' + a_{21}y')^2 + (a_{12}x' + a_{22}y')^2 \\
&= \left(a_{11}^2 + a_{12}^2\right)(x')^2 + 2(a_{11}a_{21} + a_{12}a_{22})x'y' + \left(a_{21}^2 + a_{22}^2\right)(y')^2.
\end{aligned}$$

Si recordamos que la matriz de cambio de base M es tal que

$$\begin{aligned}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} &= MM^{-1} \\ &= MM^t \\ &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22}, \end{bmatrix}\end{aligned}$$

concluimos que $a_{11}^2 + a_{12}^2 = 1 = a_{21}^2 + a_{22}^2$ y que $a_{11}a_{21} + a_{12}a_{22} = 0$, de modo que

$$\begin{aligned} f(\hat{x}) &= f(x', y') \\ &= (x')^2 + (y')^2. \end{aligned}$$

No debería causar sorpresa al lector que las expresiones de f en términos de las coordenadas (x, y) y de las coordenadas (x', y') sean análogas en este caso ya que, en términos geométricos, f asigna a cada punto $\hat{x}$ su norma al cuadrado (cantidad que se expresa de forma análoga en ambos sistemas coordenados, en virtud de que éstos son ortonormales).

De esta forma, si las coordenadas del punto $\hat{x}_0$ y del vector $\hat{u}$ en el sistema coordenado determinado por la base $\{\hat{e}_1', \hat{e}_2'\}$ están dadas por (x_0', y_0') y (u_1', u_2') respectivamente, al calcular $D_{\hat{u}} f(\hat{x}_0)$ usando la expresión de f en este sistema y repitiendo los cálculos que hicimos al principio, obtendremos nuevamente que

$$D_{\hat{u}} f(\hat{x}_0) = 2(x_0' u_1' + y_0' u_2').$$

Lo que ahora mostraremos es que los dos valores que obtuvimos para $D_{\hat{u}} f(\hat{x}_0)$ usando la expresión de f en cada uno de los sistemas coordenados, coinciden. Para ello, recordemos que las coordenadas del punto $\hat{x}_0$ y del vector $\hat{u}$ en ambos sistemas también satisfacen las identidades

$$\begin{bmatrix} x_0 & y_0 \end{bmatrix} = \begin{bmatrix} x_0' & y_0' \end{bmatrix} M$$

y

$$\begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} u_1' & u_2' \end{bmatrix} M$$

respectivamente, de tal forma que, usando la notación del producto de matrices, obtenemos que

$$\begin{aligned}
2(x_0u_1 + y_0u_2) &= 2\begin{bmatrix} x_0 & y_0 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \\
&= 2\left(\begin{bmatrix} x_0' & y_0' \end{bmatrix} M\right)\left(\begin{bmatrix} u_1' & u_2' \end{bmatrix} M\right)^t \\
&= 2\begin{bmatrix} x_0' & y_0' \end{bmatrix}(MM^t)\begin{bmatrix} u_1' \\ u_2' \end{bmatrix} \\
&= 2\begin{bmatrix} x_0' & y_0' \end{bmatrix}(MM^{-1})\begin{bmatrix} u_1' \\ u_2' \end{bmatrix} \\
&= 2\begin{bmatrix} x_0' & y_0' \end{bmatrix}\begin{bmatrix} u_1' \\ u_2' \end{bmatrix} \\
&= 2(x_0'u_1' + y_0'u_2').
\end{aligned}$$

Aun cuando lo que acabamos de hacer sólo muestra para un ejemplo específico que el valor de $D_{\hat{u}}f(\hat{x}_0)$ no depende del sistema coordenado que usemos, parte del material que vamos a desarrollar en este capítulo será útil para probar que este concepto es, en general, independiente de dichos sistemas.

Dada la similitud que existe entre el concepto de derivada direccional y el de derivada para funciones de $\mathbb{R}$ en $\mathbb{R}$, es de suponerse que muchas de las propiedades de esta última también las satisfaga la primera. Una de estas propiedades, que en el caso real resulta ser muy importante, es el hecho de que toda función que sea derivable en un punto tiene que ser continua en ese punto.

Dado que la derivada direccional $D_{\hat{u}}f(\hat{x}_0)$ sólo toma en cuenta la variación de la función sobre la recta que pasa por $\hat{x}_0$ en la dirección del vector $\hat{u}$, lo que podemos probar es que si dicha derivada direccional existe, entonces la función, restringida a dicha recta, será continua en el punto $\hat{x}_0$.

Este hecho queda plasmado en la siguiente proposición y dejamos su prueba al lector.

Proposición 4.1 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Si $D_{\hat{u}} f(\hat{x}_0)$ existe, entonces*

$$\lim_{h \to 0} f(\hat{x}_0 + h\hat{u}) = f(\hat{x}_0).$$

Otras propiedades importantes de la derivada direccional son las relacionadas con la aritmética de las funciones. Nuevamente, de forma análoga a lo que sucede con las funciones de $\mathbb{R}$ en $\mathbb{R}$, además de que la existencia de la derivada direccional se preserva bajo dicha aritmética, las fórmulas de derivación de las funciones que se obtienen al realizar esta aritmética, resultarán ser muy útiles.

Proposición 4.2 *Sean $f, g : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Si $D_{\hat{u}} f(\hat{x}_0)$ y $D_{\hat{u}} g(\hat{x}_0)$ existen, entonces:*

1. *$D_{\hat{u}}(f + g)(\hat{x}_0)$ existe y además*

$$D_{\hat{u}}(f + g)(\hat{x}_0) = D_{\hat{u}} f(\hat{x}_0) + D_{\hat{u}} g(\hat{x}_0)$$

2. *si $\alpha \in \mathbb{R}$, $D_{\hat{u}}(\alpha f(\hat{x}_0)$ existe y además*

$$D_{\hat{u}}(\alpha f)(\hat{x}_0) = \alpha D_{\hat{u}} f(\hat{x}_0)$$

3. *$D_{\hat{u}}(fg)(\hat{x}_0)$ existe y además*

$$D_{\hat{u}}(fg)(\hat{x}_0) = f(\hat{x}_0) D_{\hat{u}} g(\hat{x}_0) + g(\hat{x}_0) D_{\hat{u}} f(\hat{x}_0)$$

4. *si g es continua en $\hat{x}_0$ y $g(\hat{x}_0) \neq 0$, $D_{\hat{u}}(f/g)(\hat{x}_0)$ existe y además*

$$D_{\hat{u}}(f/g)(\hat{x}_0) = \frac{g(\hat{x}_0) D_{\hat{u}} f(\hat{x}_0) - f(\hat{x}_0) D_{\hat{u}} g(\hat{x}_0)}{g^2(\hat{x}_0)}.$$

Aun cuando la prueba de esta proposición se deja al lector, es importante hacer un comentario acerca de la hipótesis de continuidad que se pide en el inciso 4. En realidad, dicha hipótesis se incluye para seguir garantizando

la condición que nos impusimos de que las funciones con las que vamos a trabajar a partir de este capítulo estén definidas sobre un conjunto abierto. Nótese que por el inciso (a) del problema 43 del capítulo 2, y del hecho de que el dominio de g sea el abierto U, podemos asegurar que la función f/g está definida en un abierto que contiene al punto $\hat{x}_0$.

Para finalizar con las propiedades de la derivada direccional relacionadas con las operaciones entre funciones, formularemos en una proposición aparte aquella que nos habla de la composición de funciones.

El caso que va a resultar más interesante será cuando compongamos a f con una función g de $\mathbb{R}$ en $\mathbb{R}$. Esta propiedad, que en realidad es una "consecuencia" de un resultado más general conocido como "la regla de la cadena" (que probaremos más adelante), queda plasmada en la siguiente

Proposición 4.3 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$ y $(a, b) \subset \mathbb{R}$ tal que $f(U) \subset (a, b)$. Si $g : (a, b) \subset \mathbb{R} \to \mathbb{R}$ es derivable en $f(\hat{x}_0)$ y la $D_{\hat{u}} f(\hat{x}_0)$ existe, entonces la $D_{\hat{u}}(g \circ f)(\hat{x}_0)$ existe y además*

$$D_{\hat{u}}(g \circ f)(\hat{x}_0) = g'(f(\hat{x}_0)) D_{\hat{u}} f(\hat{x}_0).$$

Demostración. De acuerdo con la definición de $D_{\hat{u}}(g \circ f)(\hat{x}_0)$, debemos probar que

$$\lim_{h \to 0} \frac{(g \circ f)(\hat{x}_0 + h\hat{u}) - (g \circ f)(\hat{x}_0)}{h} = \lim_{h \to 0} \frac{g(f(\hat{x}_0 + h\hat{u})) - g(f(\hat{x}_0))}{h}$$

existe.

Como seguramente el lector recordará de la correspondiente prueba para el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$, si hacemos

$$k_h = f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0),$$

el límite anterior se podría escribir como

$$\lim_{h \to 0} \frac{g(f(\hat{x}_0 + h\hat{u})) - g(f(\hat{x}_0))}{h} = \lim_{h \to 0} \frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{h},$$

de tal forma que si $k_h \neq 0$, entonces

$$\begin{aligned}
&\lim_{h\to 0} \frac{g(f(\hat{x}_0 + h\hat{u})) - g(f(\hat{x}_0))}{h} \\
&\quad = \lim_{h\to 0} \frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{h} \\
&\quad = \lim_{h\to 0} \left[\frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{k_h} \cdot \frac{k_h}{h} \right] \\
&\quad = \lim_{h\to 0} \left[\frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{k_h} \cdot \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h} \right].
\end{aligned}$$

Si ahora notamos que, por la proposición 4.1, se tiene que $k_h \to 0$ cuando $h \to 0$ ya que

$$\begin{aligned}
\lim_{h\to 0} k_h &= \lim_{h\to 0} (f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)) \\
&= 0,
\end{aligned}$$

entonces

$$\lim_{h\to 0} \frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{k_h} = g'(f(\hat{x}_0))$$

y obtendríamos el resultado deseado.

Como seguramente el lector ya sabe, el problema con el argumento anterior es que éste sólo funciona si $k_h \neq 0$, y como también recordará, la solución a este problema está en definir una función auxiliar.

Definimos $\varphi : (-r, r) \subset \mathbb{R} \to \mathbb{R}$, con $r > 0$ tal que $B_r(\hat{x}_0) \subset U$, de la siguiente manera:

$$\varphi(h) = \begin{cases} \frac{g(f(\hat{x}_0)+k_h)-g(f(\hat{x}_0))}{k_h} & \text{si } k_h \neq 0 \\ \\ g'(f(\hat{x}_0)) & \text{si } k_h = 0 \end{cases}$$

Lo primero que es importante observar (y que es muy fácil de verificar) es que

$$\frac{g(f(\hat{x}_0) + k_h) - g(f(\hat{x}_0))}{h} = \varphi(h) \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h}$$

para toda $h \in (-r, r)$, $h \neq 0$, de tal forma que nuestro problema se reduce a demostrar que

$$\lim_{h \to 0} \varphi(h) = g'(f(\hat{x}_0)).$$

Sea entonces $\varepsilon > 0$; dado que g es derivable en $f(\hat{x}_0)$, sabemos que existe $\delta' > 0$ tal que si $|k| < \delta'$ (y $f(\hat{x}_0) + k \in (a, b)$), entonces

$$\left| \frac{g(f(\hat{x}_0) + k) - g(f(\hat{x}_0))}{k} - g'(f(\hat{x}_0)) \right| < \varepsilon. \tag{4.10}$$

Por otra parte, por la proposición 4.1 sabemos que

$$\begin{aligned} \lim_{h \to 0} k_h &= \lim_{h \to 0} (f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)) \\ &= 0 \end{aligned}$$

y por tanto existe $0 < \delta \leq r$ tal que si $|h| < \delta$, entonces $|k_h| < \delta'$. De esta forma, si $|h| < \delta$, independientemente de que $k_h \neq 0$ o $k_h = 0$, por la desigualdad 4.10 o por el valor de φ cuando $k_h = 0$, en ambos casos se tiene que

$$\left| \varphi(h) - g'(f(\hat{x}_0)) \right| < \varepsilon$$

y por lo tanto que

$$\lim_{h \to 0} \varphi(h) = g'(f(\hat{x}_0)),$$

con lo cual concluimos la prueba. ■

Terminamos esta serie de proposiciones con una en la que se establece una propiedad que bien podría interpretarse como la versión del Teorema del Valor Medio para la derivada direccional. Su formulación es la siguiente, en donde recordamos que $[\hat{a}, \hat{b}]$ representa al segmento de recta que une al punto $\hat{a}$ con el punto $\hat{b}$.

Proposición 4.4 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{a}, \hat{b} \in U$, $\hat{a} \neq \hat{b}$, tales que $\left[\hat{a}, \hat{b}\right] \subset U$ y $\hat{u} = \left(\hat{b} - \hat{a}\right) / \left\|\hat{b} - \hat{a}\right\| \in \mathbb{R}^n$. Si $D_{\hat{u}} f(\hat{x})$ existe para toda $\hat{x} \in \left[\hat{a}, \hat{b}\right]$,*

entonces existe $\xi \in \left(0, \left\|\hat{b} - \hat{a}\right\|\right)$ *tal que si* $\hat{\xi} = \hat{a} + \xi\hat{u}$, *se cumple que*

$$f(\hat{b}) - f(\hat{a}) = \left\|\hat{b} - \hat{a}\right\| D_{\hat{u}} f\left(\hat{\xi}\right).$$

Demostración. Como es de suponerse, la prueba de esta proposición se basa en el Teorema del Valor Medio para funciones de $\mathbb{R}$ en $\mathbb{R}$. Definimos $g : \left[0, \left\|\hat{b} - \hat{a}\right\|\right] \subset \mathbb{R} \to \mathbb{R}$ como $g(t) = f(\hat{a} + t\hat{u})$. Nótese que g es derivable para toda $t \in \left[0, \left\|\hat{b} - \hat{a}\right\|\right]$, puesto que

$$\begin{aligned} g'(t) &= \lim_{h\to 0} \frac{g(t+h) - g(t)}{h} \\ &= \lim_{h\to 0} \frac{f(\hat{a} + (t+h)\hat{u}) - f(\hat{a} + t\hat{u})}{h} \\ &= \lim_{h\to 0} \frac{f((\hat{a} + t\hat{u}) + h\hat{u}) - f(\hat{a} + t\hat{u})}{h} \\ &= D_{\hat{u}} f(\hat{a} + t\hat{u}), \end{aligned}$$

de tal forma que, por el Teorema del Valor Medio para funciones $\mathbb{R}$ en $\mathbb{R}$, se tiene que existe $\xi \in \left(0, \left\|\hat{b} - \hat{a}\right\|\right)$ tal que

$$\begin{aligned} f\left(\hat{b}\right) - f(\hat{a}) &= g\left(\left\|\hat{b} - \hat{a}\right\|\right) - g(0) \\ &= \left(\left\|\hat{b} - \hat{a}\right\| - 0\right) g'(\xi) \\ &= \left\|\hat{b} - \hat{a}\right\| D_{\hat{u}} f(\hat{a} + \xi\hat{u}) \\ &= \left\|\hat{b} - \hat{a}\right\| D_{\hat{u}} f\left(\hat{\xi}\right). \end{aligned}$$

■

4.2.1. Derivadas parciales

Si una función f está escrita en términos de las coordenadas $x_1, \ldots, x_n$ asociadas a un sistema de referencia determinado por una base ortonormal

$\{\hat{e}_1, \ldots, \hat{e}_n\}$ de $\mathbb{R}^n$, calcular la derivada direccional de f en la dirección de estos vectores básicos (en cualquier punto $\hat{x}$ del dominio de f) resultará más sencillo (y más importante) que en cualquier otra dirección. En efecto, si

$$\begin{aligned}\hat{x} &= x_1\hat{e}_1 + \cdots + x_n\hat{e}_n \\ &= (x_1, \ldots, x_n)\end{aligned}$$

y el valor de f en $\hat{x}$ se puede escribir en términos de las coordenadas $(x_1, \ldots, x_n)$, es decir que

$$f(\hat{x}) = f(x_1, \ldots, x_n),$$

entonces el cálculo de la derivada direccional de f en $\hat{x}$, en la dirección de cada vector $\hat{e}_i$ se traduce en lo siguiente: de la definición de la derivada direccional $D_{\hat{e}_i} f(\hat{x})$ sabemos que

$$D_{\hat{e}_i} f(\hat{x}) = \lim_{h \to 0} \frac{f(\hat{x} + h\hat{e}_i) - f(\hat{x})}{h}$$

y dado que las coordenadas del vector $\hat{x} + h\hat{e}_i$ son $(x_1, \ldots, x_i + h, \ldots, x_n)$, el límite anterior escrito en términos de coordenadas se convierte en

$$\begin{aligned}D_{\hat{e}_i} f(\hat{x}) &= \lim_{h \to 0} \frac{f(\hat{x} + h\hat{e}_i) - f(\hat{x})}{h} \\ &= \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}.\end{aligned}$$

Como seguramente el lector podrá intuir, en el límite anterior la única coordenada en la que se está teniendo un incremento h es en la $i-$ésima, mientras que en las otras coordenadas no hay cambios, es decir, permanecen fijas. En términos más informales, esto significa que para el cálculo de la derivada direccional $D_{\hat{e}_i} f(\hat{x})$ bastaría con derivar la expresión $f(x_1, \ldots, x_n)$ tomando como única variable a x_i, y considerando a las restantes $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ como si fueran constantes. En el siguiente ejemplo ilustramos este hecho de manera más clara.

Ejemplo 4.3 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ la función cuyo valor en un punto $\hat{x} \in \mathbb{R}^2$ está dado en términos de sus coordenadas (x, y) (en la base canónica, a cuyos elementos denotaremos (en este caso) por $\hat{e}_x$ y $\hat{e}_y$) por la expresión*

$$f(x, y) = 4x^5y^2.$$

Calcularemos $D_{\hat{e}_y} f(\hat{x})$ para cualquier $\hat{x} \in \mathbb{R}^2$. Si $\hat{x} = (x, y)$, de acuerdo con la definición de la derivada direccional $D_{\hat{e}_y} f(\hat{x})$, se tiene que

$$\begin{aligned} D_{\hat{e}_y} f(\hat{x}) &= \lim_{h\to 0} \frac{f(\hat{x} + h\hat{e}_y) - f(\hat{x})}{h} \\ &= \lim_{h\to 0} \frac{f(x, y + h) - f(x, y)}{h} \\ &= \lim_{h\to 0} \frac{4x^5(y + h)^2 - 4x^5y^2}{h} \\ &= \lim_{h\to 0} 4x^5 \frac{(y + h)^2 - y^2}{h} \\ &= 4x^5 \lim_{h\to 0} \frac{(y + h)^2 - y^2}{h} \\ &= 4x^5(2y) = 8x^5y. \end{aligned}$$

Como el lector habrá notado en este ejemplo, el cálculo de la derivada direccional $D_{\hat{e}_y} f(\hat{x})$ se redujo a derivar la expresión $4x^5y^2$ considerando sólo como variable a la coordenada y y tratando al resto de la expresión $(4x^5)$ como una constante. Una de las ventajas de lo anterior (¡entre otras más!) es que podemos simplificar el cálculo de las derivadas direccionales usando los métodos de derivación que aprendimos para las funciones de $\mathbb{R}$ en $\mathbb{R}$.

Por estas características, y algunas otras que veremos más adelante, las derivadas direccionales en la dirección de los vectores de una base ortonormal tienen un nombre y una notación propia, las cuales establecemos en la siguiente

Definición 4.2 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Si $x_1, \ldots, x_n$ denotan las variables (o coordenadas) determinadas por una*

base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$ *de* $\mathbb{R}^n$*, definimos la derivada parcial de* f *con respecto de la variable* x_i *en* $\hat{x}_0$*, que denotamos por* $\frac{\partial f}{\partial x_i}(\hat{x}_0)$*, como*

$$\begin{aligned}\frac{\partial f}{\partial x_i}(\hat{x}_0) &:= D_{\hat{e}_i} f(\hat{x}_0) \\ &:= \lim_{h\to 0} \frac{f(\hat{x}_0 + h\hat{e}_i) - f(\hat{x}_0)}{h}.\end{aligned}$$

Como es de suponerse, dado que cualquier derivada parcial no es más que una cierta derivada direccional, las proposiciones 4.1, 4.2, 4.3, y 4.4, tienen sus correspondientes versiones para derivadas parciales, las cuales formalizaremos a continuación (sin probar). En todas ellas supondremos que $x_1, \ldots, x_n$ son las variables (o coordenadas) determinadas por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$ de $\mathbb{R}^n$.

Proposición 4.5 *Sean* $f : U \subset \mathbb{R}^n \to \mathbb{R}$ *y* $\hat{x}_0 \in U$*. Si* $\frac{\partial f}{\partial x_i}(\hat{x}_0)$ *existe, entonces*

$$\lim_{h\to 0} f(\hat{x}_0 + h\hat{e}_i) = f(\hat{x}_0).$$

Proposición 4.6 *Sean* $f, g : U \subset \mathbb{R}^n \to \mathbb{R}$ *y* $\hat{x}_0 \in U$*. Si* $\frac{\partial f}{\partial x_i}(\hat{x}_0)$ *y* $\frac{\partial g}{\partial x_i}(\hat{x}_0)$ *existen para alguna* $i \in \{1, \ldots, n\}$*, entonces:*

1. $\frac{\partial (f+g)}{\partial x_i}(\hat{x}_0)$ *existe y además*

$$\frac{\partial (f+g)}{\partial x_i}(\hat{x}_0) = \frac{\partial f}{\partial x_i}(\hat{x}_0) + \frac{\partial g}{\partial x_i}(\hat{x}_0)$$

2. *si* $\alpha \in \mathbb{R}$, $\frac{\partial (\alpha f)}{\partial x_i}(\hat{x}_0)$ *existe y además*

$$\frac{\partial (\alpha f)}{\partial x_i}(\hat{x}_0) = \alpha \frac{\partial f}{\partial x_i}(\hat{x}_0)$$

3. $\frac{\partial (fg)}{\partial x_i}(\hat{x}_0)$ *existe y además*

$$\frac{\partial (fg)}{\partial x_i}(\hat{x}_0) = f(\hat{x}_0)\frac{\partial g}{\partial x_i}(\hat{x}_0) + g(\hat{x}_0)\frac{\partial f}{\partial x_i}(\hat{x}_0)$$

4. si g es continua en $\hat{x}_0$ y $g(\hat{x}_0) \neq 0$, $\frac{\partial(f/g)}{\partial x_i}(\hat{x}_0)$ existe y además

$$\frac{\partial(f/g)}{\partial x_i}(\hat{x}_0) = \frac{g(\hat{x}_0)\frac{\partial f}{\partial x_i}(\hat{x}_0) - f(\hat{x}_0)\frac{\partial g}{\partial x_i}(\hat{x}_0)}{g^2(\hat{x}_0)}.$$

Proposición 4.7 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $(a,b) \subset \mathbb{R}$ tal que $f(U) \subset (a,b)$. Si $g : (a,b) \subset \mathbb{R} \to \mathbb{R}$ es derivable en $f(\hat{x}_0)$ y $\frac{\partial f}{\partial x_i}(\hat{x}_0)$ existe para alguna $i \in \{1,\ldots,n\}$, entonces $\frac{\partial(g\circ f)}{\partial x_i}(\hat{x}_0)$ existe y además*

$$\frac{\partial(g \circ f)}{\partial x_i}(\hat{x}_0) = g'(f(\hat{x}_0))\frac{\partial f}{\partial x_i}(\hat{x}_0).$$

Proposición 4.8 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{a}, \hat{b} \in U$, $\hat{a} \neq \hat{b}$, tales que $\left[\hat{a}, \hat{b}\right] \subset U$ y $\hat{b} - \hat{a} = (b-a)\hat{e}_i$. Si para alguna $i \in \{1,\ldots,n\}$ se tiene que $\frac{\partial f}{\partial x_i}(\hat{x})$ existe para toda $\hat{x} \in \left[\hat{a}, \hat{b}\right]$, entonces existe $\xi \in \left(0, \left\|\hat{b} - \hat{a}\right\| = |b-a|\right)$ tal que si $\hat{\xi} = \hat{a} + \xi\hat{e}_i \in \left[\hat{a}, \hat{b}\right]$, se cumple que*

$$f(\hat{b}) - f(\hat{a}) = (b-a)\frac{\partial f}{\partial x_i}(\hat{\xi}). \tag{4.11}$$

En esta última proposición es importante hacer notar dos aspectos relevantes: uno, que la hipótesis de que $\hat{a}, \hat{b} \in U$ sean tales que $\hat{b} - \hat{a} = (b-a)\hat{e}_i$ significa que las coordenadas de $\hat{a}$ y $\hat{b}$ (en la base $\{\hat{e}_1,\ldots,\hat{e}_n\}$) sólo difieren en la $i-$ésima coordenada; y dos, que en el caso en que la cantidad $b-a$ sea negativa, entonces el vector $\hat{u}$ que se construye en la proposición 4.4 es tal que

$$\begin{aligned}\hat{u} &= \frac{\hat{b} - \hat{a}}{\left\|\hat{b} - \hat{a}\right\|} \\ &= \frac{b-a}{|b-a|}\hat{e}_i \\ &= -\hat{e}_i.\end{aligned}$$

De esta forma, de acuerdo con la conclusión de dicha proposición (y el fácil resultado que el lector probará en el problema 1 de este capítulo), se tiene que

$$\begin{aligned} f(\hat{b}) - f(\hat{a}) &= \left\| \hat{b} - \hat{a} \right\| D_{\hat{u}} f(\hat{\xi}) \\ &= |b-a| \, D_{-\hat{e}_i} f(\hat{\xi}) \\ &= -(b-a)(-D_{\hat{e}_i} f(\hat{\xi})) \\ &= (b-a) \, \frac{\partial f}{\partial x_i}(\hat{\xi}), \end{aligned}$$

lo que prueba que la identidad 4.11 se cumple independientemente de la relación (de orden) que guarden los números a y b.

4.3. La derivada global

Salvo por el caso de las derivadas direccionales, quedarnos sólo con la idea de que la derivada es una forma de medir "la razón de cambio" de una función, es algo que dificilmente nos ayudará a "extender" el concepto de derivada a funciones de $\mathbb{R}^n$ en $\mathbb{R}$. Por esta razón, en esta sección empezaremos por recordar que la derivabilidad de una función f de $\mathbb{R}$ en $\mathbb{R}$ en un punto x_0, además de proporcionarnos una medida de "la razón de cambio" de f en x_0, también nos ayuda a resolver un problema geométrico: encontrar la recta tangente a la gráfica de la función en el punto $(x_0, f(x_0))$.

Como el lector seguramente recordará, la derivada de una función f de $\mathbb{R}$ en $\mathbb{R}$ en un punto x_0 $(f'(x_0))$ tiene la propiedad de que, si tomamos la recta con esta pendiente que pasa por el punto $(x_0, f(x_0))$, esta recta es tangente a la gráfica de la función en dicho punto.

Específicamente, lo anterior se traduce en lo siguiente: la recta con pendiente $f'(x_0)$ que pasa por el punto $(x_0, f(x_0))$ es aquella cuya ecuación se puede escribir de la forma

$$y = f'(x_0)(x - x_0) + f(x_0).$$

Que esta recta sea tangente a la gráfica de f en $(x_0, f(x_0))$ significa no sólo que pasa por ese punto, sino que además se "parece" mucho a f cerca de x_0, donde eso de "parecerse mucho a f cerca de x_0" se traduce en que

$$\lim_{x\to x_0} \frac{f(x) - (f'(x_0)(x - x_0) + f(x_0))}{x - x_0} = 0,$$

es decir, que la diferencia $f(x) - (f'(x_0)(x - x_0) + f(x_0))$ se va más "rápido" a cero de lo que se va la diferencia $x - x_0$. De hecho, como seguramente el lector también recordará, la existencia de una recta con estas características garantiza la derivabilidad de la función en el punto x_0. Dado que las rectas que pasan por el punto $(x_0, f(x_0))$ tienen una ecuación de la forma

$$y = m(x - x_0) + f(x_0),$$

si una de estas rectas tiene la propiedad de que

$$\lim_{x\to x_0} \frac{f(x) - (m(x - x_0) + f(x_0))}{x - x_0} = 0, \tag{4.12}$$

entonces podemos asegurar que f es derivable en x_0 y que además $f'(x_0) = m$. En efecto, como

$$\begin{aligned} 0 &= \lim_{x\to x_0} \frac{f(x) - (m(x - x_0) + f(x_0))}{x - x_0} \\ &= \lim_{x\to x_0} \left[\frac{f(x) - f(x_0)}{x - x_0} - m \right], \end{aligned}$$

entonces

$$\begin{aligned} m &= \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0} \\ &= f'(x_0), \end{aligned}$$

lo que comprueba nuestra afirmación.

De lo anterior concluimos que la derivabilidad de una función f de $\mathbb{R}$ en $\mathbb{R}$ en un punto x_0 es equivalente a la existencia de una función de la

forma $r(x) = m(x - x_0) + f(x_0)$ que satisfaga la identidad 4.12. En realidad bastaría considerar las funciones de la forma $L(x) = mx$, puesto que la función $r(x)$ no es más que, en términos de sus gráficas, la "traslación" de $L(x)$ (que pasa por el origen) al punto $(x_0, f(x_0))$ (ver figura 4.4).

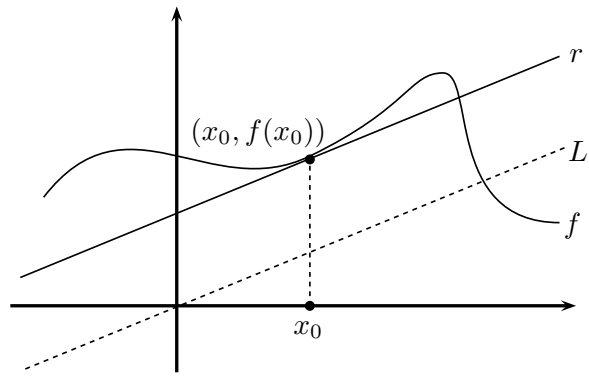


Figura 4.4: La función $r(x) = m(x - x_0) + f(x_0)$ es la "traslación" de la función $L(x) = mx$ (que pasa por el origen) al punto $(x_0, f(x_0))$.

Las funciones de la forma $L(x) = mx$ son justo el tipo de funciones de $\mathbb{R}$ en $\mathbb{R}$ conocidas con el nombre de funciones lineales (las cuales introdujimos en el problema 38 del capítulo 2 para el caso general de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$), y que son las funciones $L : \mathbb{R} \to \mathbb{R}$ que tienen las siguientes dos propiedades:

1. $L(x + y) = L(x) + L(y)$ para todas $x, y \in \mathbb{R}$, y
2. $L(\alpha x) = \alpha L(x)$ para todas $\alpha, x \in \mathbb{R}$.

Es un problema sencillo mostrar que $L : \mathbb{R} \to \mathbb{R}$ *es una función lineal si y sólo si existe* $m \in \mathbb{R}$ *tal que* $L(x) = mx$ *para toda* $x \in \mathbb{R}$ (en ambas implicaciones se deduce que $m = L(1)$).

Como seguramente el lector ya estará imaginando, la discusión anterior nos sugiere la forma en que podemos dar la definición de derivada (global) de una función de $\mathbb{R}^n$ en $\mathbb{R}$. En términos generales, que una función f de

este tipo sea derivable en un punto $\hat{x}_0$ de su dominio significará que existe una función lineal L de $\mathbb{R}^n$ en $\mathbb{R}$ tal que al "trasladarla" para que su valor en $\hat{x}_0$ sea $f(\hat{x}_0)$, esta función lineal "trasladada" $(L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))$, que recibe el nombre de *función (o transformación) afín* se "parece" mucho a la función f "cerca" de $\hat{x}_0$.

Específicamente, nuestra definición de derivabilidad será la siguiente.

Definición 4.3 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Decimos que f es derivable en $\hat{x}_0$ si existe una función lineal $L : \mathbb{R}^n \to \mathbb{R}$ tal que*

$$\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - (L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} = 0. \tag{4.13}$$

Sin duda que después de esta definición, sería conveniente dar un ejemplo que la ilustrara. Sin embargo, es importante hacer notar que hasta ahora no tenemos ninguna herramienta que nos permita, dada una función (expresada en algún sistema coordenado), proponer (o intuir) cuál debería de ser la función lineal que satisfaga la condición que se le pide en esta definición.

Por esta razón, primero nos enfocaremos en dar algunas propiedades relacionadas con este concepto de derivada que, entre otras cosas, nos permitirán darnos una idea de cómo encontrar la ya famosa función lineal L.

De hecho, la primera proposición que probaremos es justo la que nos asegura que sólo puede haber una función lineal que satisfaga la propiedad de la definición 4.3.

Proposición 4.9 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $L, \tilde{L} : \mathbb{R}^n \to \mathbb{R}$ funciones lineales. Si L y $\tilde{L}$ satisfacen la condición 4.13 de la definición 4.3, entonces L y $\tilde{L}$ son iguales.*

Demostración. Dado que por hipótesis sabemos que

$$\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - (L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} = 0$$

y

$$\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - (\tilde{L}(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} = 0,$$

restando ambos límites concluimos que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{\tilde{L}(\hat{x}-\hat{x}_0) - L(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|} = \lim_{\hat{x}\to\hat{x}_0} \frac{(\tilde{L}-L)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|} = 0,$$

de tal forma que si hacemos $\hat{L} = \tilde{L} - L$ y $\hat{h} = \hat{x} - \hat{x}_0$, entonces $\hat{L}$ es una función lineal para la cual se satisface que

$$\lim_{\hat{h}\to\hat{0}} \frac{\hat{L}(\hat{h})}{\left\|\hat{h}\right\|} = 0.$$

Por los problemas 33 y 38 del capítulo 2, se tiene que $\hat{L}$ es la constante cero, o lo que es lo mismo, que L y $\tilde{L}$ son iguales. ■

Con base en esta proposición ya podemos completar la definición 4.3 de la siguiente forma: si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es derivable en $\hat{x}_0 \in U$, a la función lineal que satisface la condición 4.13, que por la proposición anterior sabemos que es única, la denotaremos por $Df(\hat{x}_0)$ y diremos que es *la derivada de* f *en* $\hat{x}_0$. De esta forma, *la derivada de una función de* $\mathbb{R}^n$ *en* $\mathbb{R}$ *es una función lineal.*

Más adelante veremos que toda funcion lineal de $\mathbb{R}^n$ en $\mathbb{R}$ se puede representar (dependiendo del sistema coordenado que se elija) por una matriz de $1 \times n$ o por un vector en $\mathbb{R}^n$, de tal forma que en algunas ocasiones diremos (abusando ciertamente del lenguaje) que la derivada de este tipo de funciones también es (o dicho de manera más correcta, que se puede representar por) una de estas matrices (o uno de estos vectores).

Justamente, con el objetivo de poder "calcular" (o conocer) la derivada de una función en un punto $\hat{x}_0$, el siguiente resultado que veremos establece una importante relación entre la función lineal $Df(\hat{x}_0)$ (que es la derivada en el punto $\hat{x}_0$) y las derivadas direccionales en dicho punto.

En términos generales, la siguiente proposición que probaremos nos asegura que si una función es derivable en un punto $\hat{x}_0$, entonces la derivada direccional en $\hat{x}_0$, en la dirección de cualquier vector $\hat{u}$, también existe y

además el valor de dicha derivada direccional se obtiene "evaluando" la función lineal $Df(\hat{x}_0)$ en el vector $\hat{u}$.

Proposición 4.10 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Si f es derivable en $\hat{x}_0$, entonces la derivada direccional de f en $\hat{x}_0$ en la dirección de $\hat{u}$ también existe y además*

$$D_{\hat{u}} f(\hat{x}_0) = Df(\hat{x}_0)(\hat{u}).$$

Demostración. De la definición 4.3 sabemos que

$$\begin{aligned}
&\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} \\
&\quad = \lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \\
&\quad = 0
\end{aligned}$$

de tal forma que si en particular tomamos $\hat{x} = \hat{x}_0 + h\hat{u}$, tenemos entonces que $\hat{x} \to \hat{x}_0$ si $h \to 0$ y por lo tanto se tendrá que

$$\begin{aligned}
0 &= \lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \\
&= \lim_{\hat{x}\to\hat{x}_0} \frac{|f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)|}{\|\hat{x} - \hat{x}_0\|} \\
&= \lim_{h\to 0} \frac{|f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0) - Df(\hat{x}_0)(h\hat{u})|}{\|h\hat{u}\|} \\
&= \lim_{h\to 0} \frac{|f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0) - Df(\hat{x}_0)(h\hat{u})|}{|h|} \\
&= \lim_{h\to 0} \left| \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0) - hDf(\hat{x}_0)(\hat{u})}{h} \right| \\
&= \lim_{h\to 0} \left| \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h} - Df(\hat{x}_0)(\hat{u}) \right|,
\end{aligned}$$

lo cual es equivalente a que

$$\lim_{h\to 0} \frac{f(\hat{x}_0 + h\hat{u}) - f(\hat{x}_0)}{h} = Df(\hat{x}_0)(\hat{u}).$$

De aquí concluimos que la derivada direccional de f en $\hat{x}_0$ en la dirección de $\hat{u}$ existe y además que

$$D_{\hat{u}} f(\hat{x}_0) = Df(\hat{x}_0)(\hat{u}).$$

■

Para abordar el problema de calcular la derivada de una función es necesario establecer algunos hechos importantes acerca de las funciones lineales, lo cual haremos a continuación.

Es fácil ver que si $L : \mathbb{R}^n \to \mathbb{R}$ es una función lineal, para conocer el valor de L en cualquier punto $\hat{x} \in \mathbb{R}^n$, es suficiente con saber los valores de L en los elementos de una base de $\mathbb{R}^n$. En efecto, nótese que si $\hat{x}$ tiene coordenadas $(x_1, \ldots, x_n)$ en un sistema coordenado determinado por una base $\beta = \{\hat{v}_1, \ldots, \hat{v}_n\}$, esto significa que

$$\hat{x} = x_1\hat{v}_1 + \cdots + x_n\hat{v}_n,$$

de tal forma que, por la "linealidad" de L se tiene que

$$\begin{aligned} L(\hat{x}) &= L(x_1\hat{v}_1 + \cdots + x_n\hat{v}_n) \\ &= x_1L(\hat{v}_1) + \cdots + x_nL(\hat{v}_n). \end{aligned}$$

Esto confirma que para saber el valor de L en $\hat{x}$ es suficiente con conocer las coordenadas de $\hat{x}$ en una base $\beta = \{\hat{v}_1, \ldots, \hat{v}_n\}$ y los valores de L en los elemento de esta base.

Si denotamos por $a_{1i} = L(\hat{v}_i)$, se suele construir la matriz de $1 \times n$

$$M_\beta = \begin{bmatrix} a_{11} & \cdots & a_{1n} \end{bmatrix}$$

y decir que la matriz M_β "representa" a la función lineal L en la base $\beta = \{\hat{v}_1, \ldots, \hat{v}_n\}$.

De hecho, esta representación matricial y las operaciones entre matrices son útiles para expresar a $L(\hat{x})$ como

$$\begin{aligned} L(\hat{x}) &= L(x_1, \ldots, x_n) \\ &= \begin{bmatrix} a_{11} & \cdots & a_{1n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= M_\beta \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \end{aligned}$$

En el caso particular de funciones lineales de $\mathbb{R}^n$ en $\mathbb{R}$, también podemos "representar" a L por el vector $(L(\hat{v}_1), \ldots, L(\hat{v}_n))$ y usar el producto punto para expresar a $L(\hat{x})$ como

$$\begin{aligned} L(\hat{x}) &= L(x_1, \ldots, x_n) \\ &= (L(\hat{v}_1), \ldots, L(\hat{v}_n)) \cdot (x_1, \ldots, x_n). \end{aligned} \tag{4.14}$$

Con base en lo anterior, si una función f está expresada en términos de las coordenadas $x_1, \ldots, x_n$ determinadas por alguna base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$ de $\mathbb{R}^n$, para encontrar la matriz que representa (en la misma base $\{\hat{e}_1, \ldots, \hat{e}_n\}$) a la derivada de f en un punto $\hat{x}_0$ de su dominio, es decir $Df(\hat{x}_0)$, bastaría con saber el valor $Df(\hat{x}_0)(\hat{e}_i)$ (para cada $i \in \{1, \ldots, n\}$). De acuerdo con lo probado en la proposición 4.10, se tiene que

$$Df(\hat{x}_0)(\hat{e}_i) = D_{\hat{e}_i} f(\hat{x}_0).$$

Si ahora recordamos, de acuerdo con la definición 4.2, que

$$D_{\hat{e}_i} f(\hat{x}_0) = \frac{\partial f}{\partial x_i}(\hat{x}_0),$$

entonces concluimos que para "calcular" (o "conocer") a $Df(\hat{x}_0)$ basta con calcular las derivadas parciales de f en $\hat{x}_0$.

En cuanto a la discusión previa, es importante hacer notar que la mera existencia de las derivadas parciales de una función f en un punto $\hat{x}_0$ de su dominio no garantizan que la función sea derivable en dicho punto. Lo que en el fondo está sucediendo, es que en la proposición 4.10 se establece que la existencia de las derivadas direccionales (incluyendo las derivadas parciales) es una condición (o consecuencia) necesaria de la derivabilidad de f en $\hat{x}_0$, pero lo recíproco no es cierto (más adelante daremos un ejemplo que ilustra esta afirmación).

Sin embargo, la importancia de la discusión anterior está en que se muestra con toda precisión cuál es la única función lineal que podría ser la derivada de f en $\hat{x}_0$, lo que sin duda es un hecho de gran valor. Una vez aclarado lo anterior, pasamos a dar el siguiente

Ejemplo 4.4 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ dada en términos de las coordenadas de la base canónica por la expresión*

$$f(x, y) = x^2 + y^2 - 1.$$

Mostraremos que f es derivable en cualquier punto $\hat{x}_0 = (x_0, y_0) \in \mathbb{R}^2$ y calcularemos $Df(\hat{x}_0)$.

De acuerdo con lo visto anteriormente, hay que calcular las derivadas parciales de f con respecto de las variables x y y,

$$\frac{\partial f}{\partial x}(x, y) = 2x \qquad y \qquad \frac{\partial f}{\partial y}(x, y) = 2y,$$

evaluarlas en $\hat{x}_0 = (x_0, y_0)$

$$\frac{\partial f}{\partial x}(x_0, y_0) = 2x_0 \qquad y \qquad \frac{\partial f}{\partial y}(x_0, y_0) = 2y_0$$

y mostrar que la función lineal (expresada en el mismo sistema coordenado)

$$\begin{aligned} L(\hat{x}) &= L(x,y) \\ &= \begin{pmatrix} \frac{\partial f}{\partial x}(x_0,y_0) & \frac{\partial f}{\partial y}(x_0,y_0) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} 2x_0 & 2y_0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= 2x_0x + 2y_0y \end{aligned}$$

satisface la condición 4.13 de la definición 4.3.

En efecto, sustituyendo f y L en dicha expresión, tenemos que

$$\begin{aligned} &\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - (L(\hat{x}-\hat{x}_0) + f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|} \\ &= \lim_{\hat{x}\to\hat{x}_0} \frac{x^2+y^2-1-(2x_0(x-x_0)+2y_0(y-y_0)+x_0^2+y_0^2-1)}{\sqrt{(x-x_0)^2+(y-y_0)^2}} \\ &= \lim_{\hat{x}\to\hat{x}_0} \frac{(x^2-2x_0x+x_0^2)+(y^2-2y_0y+y_0^2)}{\sqrt{(x-x_0)^2+(y-y_0)^2}} \\ &= \lim_{\hat{x}\to\hat{x}_0} \frac{(x-x_0)^2+(y-y_0)^2}{\sqrt{(x-x_0)^2+(y-y_0)^2}} \\ &= \lim_{\hat{x}\to\hat{x}_0} \sqrt{(x-x_0)^2+(y-y_0)^2} \\ &= 0, \end{aligned}$$

de donde concluimos que f es derivable en $\hat{x}_0$ y que además $Df(\hat{x}_0)$ es la función lineal "representada" (o "asociada"), en el mismo sistema coordenado en que está expresada f, por la matriz

$$\begin{bmatrix} \frac{\partial f}{\partial x}(x_0,y_0) & \frac{\partial f}{\partial y}(x_0,y_0) \end{bmatrix} = \begin{bmatrix} 2x_0 & 2y_0 \end{bmatrix}.$$

Es decir, que

$$\begin{aligned} Df(\hat{x}_0)(x,y) &= \begin{bmatrix} 2x_0 & 2y_0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= 2x_0x + 2y_0y. \end{aligned}$$

Para aprovechar el ejemplo anterior, en el cual podemos "ver" la gráfica de la función f, y confirmar la estrecha relación que existe entre el concepto de derivada y el concepto de tangencia, observemos que la gráfica de la función lineal

$$Df(\hat{x}_0)(x, y) = 2x_0x + 2y_0y$$

es un plano (que contiene al origen) y que la gráfica de la función

$$\begin{aligned} L(x, y) &= Df(\hat{x}_0)(x - x_0, y - y_0) + f(x_0, y_0) \\ &= 2x_0(x - x_0) + 2y_0(y - y_0) + x_0^2 + y_0^2 - 1 \end{aligned}$$

es un plano ("trasladado" al punto $(x_0, y_0, x_0^2 + y_0^2 - 1)$) que se ve "tangente" a la gráfica de la función $f(x, y) = x^2 + y^2 - 1$ en este mismo punto (ver figura 4.5).

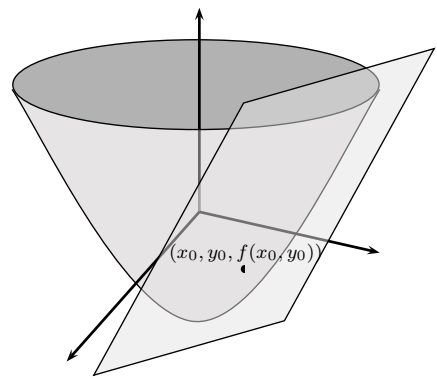


Figura 4.5: El plano "tangente" a la gráfica de la función $f(x, y) = x^2 + y^2 - 1$ en el punto $(x_0, y_0, f(x_0, y_0))$.

Con base en las observaciones anteriores, damos la siguiente

Definición 4.4 *Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ derivable en el punto $\hat{x}_0 = (x_0, y_0) \in U$. Decimos que el plano que tiene como ecuación*

$$z = \frac{\partial f}{\partial x}(x_0, y_0)\,(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)\,(y - y_0) + f(x_0, y_0) \tag{4.15}$$

es el plano tangente a la gráfica de f en el punto $(x_0, y_0, f(x_0, y_0))$.

Nótese que la ecuación 4.15 se puede escribir como

$$\left(-\frac{\partial f}{\partial x}(x_0, y_0), -\frac{\partial f}{\partial y}(x_0, y_0), 1\right) \cdot (x - x_0, y - y_0, z - f(x_0, y_0)) = 0 \quad (4.16)$$

de tal forma que el vector

$$\left(-\frac{\partial f}{\partial x}(x_0, y_0), -\frac{\partial f}{\partial y}(x_0, y_0), 1\right)$$

es un vector normal a este plano tangente.

De aquí en adelante escribiremos, sin duda abusando de la notación, que la función lineal $Df(\hat{x}_0)$ es igual a la matriz (de $1 \times n$) que la representa en el sistema de referencia que se esté usando, y en cuyas coordenadas está expresada f. Sin embargo, no hay que olvidar que esta matriz depende del sistema de referencia que se esté usando. Parte de lo que haremos a continuación será mostrar la relación que existe entre las matrices asociadas a la función lineal $Df(\hat{x}_0)$ en dos sistemas de referencia (ortonormales) diferentes, y cómo se puede obtener una a partir de la otra.

Una vez dicho lo anterior, si $x_1, \ldots, x_n$ denotan a las "variables" determinadas por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$, entonces escribiremos que

$$Df(\hat{x}_0) = \left[\begin{array}{ccc} \frac{\partial f}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\hat{x}_0) \end{array}\right], \quad (4.17)$$

de tal forma que si $\hat{u} \in \mathbb{R}^n$ tiene coordenadas $(u_1, \ldots, u_n)$ en la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$, se tiene que

$$\begin{aligned} Df(\hat{x}_0)(\hat{u}) &= \left[\begin{array}{ccc} \frac{\partial f}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\hat{x}_0) \end{array}\right] \left[\begin{array}{c} u_1 \\ \vdots \\ u_n \end{array}\right] \\ &= u_1 \frac{\partial f}{\partial x_1}(\hat{x}_0) + \cdots + u_n \frac{\partial f}{\partial x_n}(\hat{x}_0). \end{aligned}$$

Así pues, si ahora $x'_1, \ldots, x'_n$ denotan a las "variables" determinadas por otra base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ de $\mathbb{R}^n$, y se tiene que

$$\begin{aligned}\hat{e}'_i &= \left(a_1^{(i)}, \ldots, a_n^{(i)}\right) \\ &= a_1^{(i)}\hat{e}_1 + \cdots + a_n^{(i)}\hat{e}_n\end{aligned}$$

para cada $i \in \{1, \ldots, n\}$, entonces

$$\begin{aligned}\frac{\partial f}{\partial x'_i}(\hat{x}_0) &= D_{\hat{e}'_i} f(\hat{x}_0) \\ &= Df(\hat{x}_0)\left(\hat{e}'_i\right) \\ &= a_1^{(i)}\frac{\partial f}{\partial x_1}(\hat{x}_0) + \cdots + a_n^{(i)}\frac{\partial f}{\partial x_n}(\hat{x}_0).\end{aligned}$$

Esto establece la forma de obtener las derivadas parciales de f, con respecto a las variables (o coordenadas) $x'_1, \ldots, x'_n$, en términos de las variables $x_1, \ldots, x_n$. Nótese que esta última identidad también se puede obtener como una cosecuencia de la proposición 4.10, puesto que, como $\hat{e}'_i$ es un vector de norma uno, y

$$\hat{e}'_i = a_1^{(i)}\hat{e}_1 + \cdots + a_n^{(i)}\hat{e}_n,$$

entonces

$$\begin{aligned}\frac{\partial f}{\partial x'_i}(\hat{x}_0) &= D_{\hat{e}'_i} f(\hat{x}_0) \\ &= Df(\hat{x}_0)\left(\hat{e}'_i\right) \\ &= Df(\hat{x}_0)\left(a_1^{(i)}\hat{e}_1 + \cdots + a_n^{(i)}\hat{e}_n\right) \\ &= a_1^{(i)}Df(\hat{x}_0)\left(\hat{e}_1\right) + \cdots + a_n^{(i)}Df(\hat{x}_0)\left(\hat{e}_n\right) \\ &= a_1^{(i)}\frac{\partial f}{\partial x_1}(\hat{x}_0) + \cdots + a_n^{(i)}\frac{\partial f}{\partial x_n}(\hat{x}_0).\end{aligned}$$

Para terminar de encontrar la relación que existe entre las matrices asociadas a la derivada $Df(\hat{x}_0)$ en dos sistemas de referencia (ortonormales)

diferentes, recordemos que en la base $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ esta función lineal estará representada por la matriz

$$\left[\begin{array}{ccc} \frac{\partial f}{\partial x'_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x'_n}(\hat{x}_0) \end{array} \right]$$

de modo que las matrices (de $1 \times n$) que representan a la función lineal $Df(\hat{x}_0)$ (en las correspondientes bases $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ y $\{\hat{e}_1, \ldots, \hat{e}_n\}$) están relacionadas por la identidad

$$\begin{aligned} &\left[\begin{array}{ccc} \frac{\partial f}{\partial x'_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x'_n}(\hat{x}_0) \end{array} \right] \\ &\qquad = \left[\begin{array}{ccc} \frac{\partial f}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\hat{x}_0) \end{array} \right] \left[\begin{array}{ccc} a_1^{(1)} & \cdots & a_1^{(n)} \\ \vdots & \ddots & \vdots \\ a_n^{(1)} & \cdots & a_n^{(n)} \end{array} \right] \end{aligned} \tag{4.18}$$

o por la identidad

$$\begin{aligned} &\left[\begin{array}{ccc} \frac{\partial f}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\hat{x}_0) \end{array} \right] \\ &\qquad = \left[\begin{array}{ccc} \frac{\partial f}{\partial x'_1}(\hat{x}_0) & \cdots & \frac{\partial f}{\partial x'_n}(\hat{x}_0) \end{array} \right] \left[\begin{array}{ccc} a_1^{(1)} & \cdots & a_n^{(1)} \\ \vdots & \ddots & \vdots \\ a_1^{(n)} & \cdots & a_n^{(n)} \end{array} \right]. \end{aligned} \tag{4.19}$$

Esta expresiones, sin duda alguna, harán que el lector recuerde las identidades 4.4 y 4.7, que son las que obtuvimos cuando analizamos el problema relacionado con el "cambio de coordenadas" (de un mismo vector $\hat{x}$) determinadas por dos bases ortonormales de $\mathbb{R}^n$.

4.3.1. El gradiente

Además de sus implicaciones prácticas para el cálculo explícito de la derivada de una función, la proposición 4.10 tiene otra consecuencia importante, que resultará muy útil para conocer el comportamiento de una función en la vecindad de un punto en el cual sea derivable.

Como asegura la proposición 4.10, si f es derivable en un punto $\hat{x}_0$, entonces la derivada direccional de f en $\hat{x}_0$ en la dirección del vector (unitario) $\hat{u}$ se obtiene evaluando la función lineal $Df(\hat{x}_0)$ (la derivada de f en $\hat{x}_0$) en $\hat{u}$. Si ahora recordamos que toda función lineal es continua (problema 64 del capítulo 2) y que el conjunto

$$S^{n-1} = \{\hat{u} \in \mathbb{R}^n \mid \|\hat{u}\| = 1\}$$

es cerrado (problema 50 del capítulo 2) y claramente acotado, por el corolario 2.3 del capítulo 2 sabemos que deben existir $\hat{u}_1$ y $\hat{u}_2$ en S^{n-1} tales que

$$Df(\hat{x}_0)(\hat{u}_1) \leq Df(\hat{x}_0)(\hat{u}) \leq Df(\hat{x}_0)(\hat{u}_2)$$

para toda $\hat{u} \in S^{n-1}$, o equivalentemente, que

$$D_{\hat{u}_1} f(\hat{x}_0) \leq D_{\hat{u}} f(\hat{x}_0) \leq D_{\hat{u}_2} f(\hat{x}_0)$$

para toda $\hat{u} \in S^{n-1}$.

Es decir, en términos más coloquiales, siempre existen direcciones $\hat{u}_1$ y $\hat{u}_2$, una en la que la razón de cambio de la función f es mínima ($\hat{u}_1$), y otra en la que la razón de cambio de la función f es máxima ($\hat{u}_2$).

De hecho, lo siguiente que probaremos es que si la $Df(\hat{x}_0)$ no es la función lineal constante cero (en cuyo caso todas las derivadas direccionales valen 0), entonces $\hat{u}_1$ y $\hat{u}_2$ son únicos y además $\hat{u}_1 = -\hat{u}_2$.

Proposición 4.11 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Si f es derivable en $\hat{x}_0$, entonces existe $\hat{u}_0 \in S^{n-1}$ tal que*

$$D_{-\hat{u}_0} f(\hat{x}_0) \leq D_{\hat{u}} f(\hat{x}_0) \leq D_{\hat{u}_0} f(\hat{x}_0)$$

para toda $\hat{u} \in S^{n-1}$. Si la derivada $Df(\hat{x}_0)$ no es la constante cero, entonces $\hat{u}_0$ es único.

Demostración. Como mencionamos anteriormente, dado que toda función lineal es continua y que el conjunto S^{n-1} es cerrado y acotado, por el corolario 2.3 del capítulo 2 sabemos que existe $\hat{u}_0 \in S^{n-1}$ tal que

$$D_{\hat{u}} f(\hat{x}_0) \leq D_{\hat{u}_0} f(\hat{x}_0)$$

para toda $\hat{u} \in S^{n-1}$.

Por otra parte, si $\hat{u} \in S^{n-1}$, entonces $-\hat{u} \in S^{n-1}$, de modo que

$$D_{-\hat{u}} f(\hat{x}_0) \leq D_{\hat{u}_0} f(\hat{x}_0).$$

Por tanto, se tiene que

$$\begin{aligned} D_{-\hat{u}_0} f(\hat{x}_0) &= Df(\hat{x}_0)(-\hat{u}_0) \\ &= -Df(\hat{x}_0)(\hat{u}_0) \\ &= -D_{\hat{u}_0} f(\hat{x}_0) \\ &\leq -D_{-\hat{u}} f(\hat{x}_0) \\ &= -Df(\hat{x}_0)(-\hat{u}) \\ &= D_{\hat{u}} f(\hat{x}_0), \end{aligned}$$

de donde obtenemos la otra desigualdad.

Para probar la unicidad de $\hat{u}_0$, supongamos ahora que $Df(\hat{x}_0)$ no es la constante cero (de modo que $Df(\hat{x}_0)(\hat{u}_0) > 0$) y que existe otro $\hat{u}_0' \in S^{n-1}$, $\hat{u}_0' \neq \hat{u}_0$, tal que $Df(\hat{x}_0)(\hat{u}_0') = Df(\hat{x}_0)(\hat{u}_0)$.

Primero notemos que si $\hat{u}_0' = -\hat{u}_0$, entonces $\hat{u}_0' + \hat{u}_0 = \hat{0}$ y por lo tanto

$$\begin{aligned} 0 &= Df(\hat{x}_0)(\hat{0}) \\ &= Df(\hat{x}_0)(\hat{u}_0' + \hat{u}_0) \\ &= Df(\hat{x}_0)(\hat{u}_0') + Df(\hat{x}_0)(\hat{u}_0) \\ &= 2Df(\hat{x}_0)(\hat{u}_0), \end{aligned}$$

de donde se tendría que $Df(\hat{x}_0)(\hat{u}_0) = 0$. Por tanto, $\hat{u}_0' + \hat{u}_0 \neq \hat{0}$, de modo que si tomamos

$$\tilde{u} = \frac{\hat{u}_0' + \hat{u}_0}{\|\hat{u}_0' + \hat{u}_0\|} \in S^{n-1},$$

entonces

$$\begin{aligned} Df(\hat{x}_0)(\tilde{u}) &= \frac{1}{\|\hat{u}_0' + \hat{u}_0\|}(Df(\hat{x}_0)(\hat{u}_0') + Df(\hat{x}_0)(\hat{u}_0)) \\ &= \frac{2}{\|\hat{u}_0' + \hat{u}_0\|} Df(\hat{x}_0)(\hat{u}_0). \end{aligned}$$

Ahora, dado que $\hat{u}_0', \hat{u}_0 \in S^{n-1}$ y que $\hat{u}_0'$ es diferente de $\hat{u}_0$ y $-\hat{u}_0$, por el inciso (b) del problema 9 del capítulo 1 se tiene que

$$\begin{aligned}\left\|\hat{u}_0' + \hat{u}_0\right\| &< \left\|\hat{u}_0'\right\| + \|\hat{u}_0\| \\ &= 2,\end{aligned}$$

de donde

$$1 < \frac{2}{\|\hat{u}_0' + \hat{u}_0\|}.$$

Por lo tanto

$$\begin{aligned}Df(\hat{x}_0)(\hat{u}_0) &< \frac{2}{\|\hat{u}_0' + \hat{u}_0\|} Df(\hat{x}_0)(\hat{u}_0) \\ &= Df(\hat{x}_0)(\tilde{u}),\end{aligned}$$

lo cual contradice la propiedad de que $Df(\hat{x}_0)(\hat{u}_0)$ es el valor máximo de $Df(\hat{x}_0)$ sobre el conjunto S^{n-1}. Esta última contradicción concluye la prueba de que $\hat{u}_0$ es único. ■

Con base en el resultado anterior, podemos dar la siguiente

Definición 4.5 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Si f es derivable en $\hat{x}_0$, definimos el gradiente de f en $\hat{x}_0$, que denotamos por $\nabla f(\hat{x}_0)$, como el vector dado por*

$$\nabla f(\hat{x}_0) := D_{\hat{u}_0} f(\hat{x}_0)\hat{u}_0,$$

en donde $\hat{u}_0 \in \mathbb{R}^n$ es el vector unitario cuya existencia se probó en la proposición 4.11.

Como seguramente el lector ya se estará preguntando, el siguiente problema que abordaremos será el de calcular esta dirección $\hat{u}_0$ en la cual la derivada direccional alcanza su máximo valor. Como siempre, este cálculo depende de que se tenga una expresión explícita de la función en cuestión (en algún sistema coordenado), y de la forma en que se represente a su derivada. Justo en la solución de este problema es que echaremos mano de la otra

forma en que podemos representar a una función lineal de $\mathbb{R}^n$ en $\mathbb{R}$. Como se recordará, además de usar el lenguaje de matrices para representar a estas funciones lineales, también podemos usar un vector y el producto punto en $\mathbb{R}^n$, como lo escribimos en la expresión 4.14. Con base en esta expresión, tenemos entonces que, si nuestra función f está expresada en términos de las coordenadas $x_1, \ldots, x_n$ (determinadas por la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$) y el vector (unitario) $\hat{u}$ tiene coordenadas $(u_1, \ldots, u_n)$, entonces podemos escribir que

$$\begin{aligned} D_{\hat{u}} f(\hat{x}_0) &= Df(\hat{x}_0)(\hat{u}) \\ &= (Df(\hat{x}_0)(\hat{e}_1), \ldots, Df(\hat{x}_0)(\hat{e}_n)) \cdot (u_1, \ldots, u_n) \\ &= (D_{\hat{e}_1} f(\hat{x}_0), \ldots, D_{\hat{e}_n} f(\hat{x}_0)) \cdot (u_1, \ldots, u_n) \\ &= \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right) \cdot (u_1, \ldots, u_n). \end{aligned} \tag{4.20}$$

Si ahora recordamos que

$$\begin{aligned} &\left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right) \cdot (u_1, \ldots, u_n) \\ &\qquad = \left\| \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right) \right\| \|(u_1, \ldots, u_n)\| \cos(\theta) \\ &\qquad = \left\| \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right) \right\| \cos(\theta) \end{aligned}$$

(identidad 1.7 del capítulo 1), en donde θ es el ángulo formado por estos vectores, concluimos que el máximo valor de $D_{\hat{u}} f(\hat{x}_0) = Df(\hat{x}_0)(\hat{u})$ (suponiendo que $Df(\hat{x}_0)$ no es la constante cero) se alcanza cuando $\cos(\theta) = 1$, es decir si $\theta = 0$, lo que significa que el vector $\hat{u}_0$ de la proposición anterior estará expresado por

$$\hat{u}_0 = \frac{\left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right)}{\left\| \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0) \right) \right\|}. \tag{4.21}$$

De esta forma, siempre que tomemos un sistema de coordenadas x_1,. . . ,x_n

(inducido por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$), el vector con coordenadas

$$\left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0)\right) \tag{4.22}$$

es el gradiente de f en $\hat{x}_0$, es decir

$$\nabla f(\hat{x}_0) = \left(\frac{\partial f}{\partial x_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\hat{x}_0)\right). \tag{4.23}$$

Es importante insistir en que, si tomamos otro sistema de coordenadas $x'_1, \ldots, x'_n$ (inducido por otra base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$), el vector con coordenadas

$$\left(\frac{\partial f}{\partial x'_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x'_n}(\hat{x}_0)\right)$$

nuevamente es tal que

$$\hat{u}_0 = \frac{\left(\frac{\partial f}{\partial x'_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x'_n}(\hat{x}_0)\right)}{\left\|\left(\frac{\partial f}{\partial x'_1}(\hat{x}_0), \ldots, \frac{\partial f}{\partial x'_n}(\hat{x}_0)\right)\right\|},$$

en donde $\hat{u}_0$ es el multicitado vector cuya existencia probamos en la proposición 4.11 (prueba que por cierto, hicimos sin recurrir a un sistema coordenado específico).

En virtud de lo anterior, dado un sistema de coordenadas específico $x_1, \ldots, x_n$, las coordenadas del gradiente de f en $\hat{x}_0$ ($\nabla f(\hat{x}_0)$) estarán dadas por 4.23, y para obtener sus coordenadas en otro sistema $x'_1, \ldots, x'_n$, simplemente habrá que recurrir a la identidad 4.18.

Desde un punto de vista más práctico, el vector gradiente de f en $\hat{x}_0$ tiene dos usos importantes: uno, para representar (y expresar) a la derivada de f en $\hat{x}_0$ (identidad 4.20), es decir que

$$Df(\hat{x}_0)(\hat{x}) = \nabla f(\hat{x}_0) \cdot \hat{x}$$

para toda $\hat{x} \in \mathbb{R}^n$; y dos, para designar la dirección en la que f tiene la máxima razón de cambio, es decir, la dirección en la que su derivada direccional (en $\hat{x}_0$) alcanza su máximo valor (identidad 4.21). Con base en lo anterior, lo siguiente que haremos será ilustrar con un ejemplo estos dos usos.

Ejemplo 4.5 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ que está dada por la expresión*

$$f(x, y) = 8 - x^2 - y^2 + x - y$$

y $\hat{x}_0 = (2, 1)$.

Tenemos entonces que

$$\frac{\partial f}{\partial x}(x, y) = -2x + 1$$

y

$$\frac{\partial f}{\partial y}(x, y) = -2y - 1,$$

de modo que

$$\begin{aligned} \nabla f(2, 1) &= \left(\frac{\partial f}{\partial x}(2, 1), \frac{\partial f}{\partial y}(2, 1) \right) \\ &= (-3, -3)\,. \end{aligned}$$

Por tanto, la derivada de f en $(2, 1)$ $(Df(2, 1))$ es la función lineal dada por la expresión

$$\begin{aligned} Df(2, 1)(x, y) &= \nabla f(2, 1) \cdot (x, y) \\ &= (-3, -3) \cdot (x, y) \\ &= -3x - 3y, \end{aligned}$$

de tal forma que la función afín

$$Df(2, 1)(x - 2, y - 1) + f(2, 1) = -3(x - 2) - 3(y - 1) + 4$$

es la que debe satisfacer la condición 4.13 de la definción 4.3.

En efecto, se tiene que

$$\begin{aligned}
&\lim_{(x,y)\to(2,1)} \frac{f(x,y)-(Df(2,1)(x-2,y-1)+f(2,1))}{\|(x,y)-(2,1)\|}\\
&= \lim_{(x,y)\to(2,1)} \frac{8-x^2-y^2+x-y-(-3(x-2)-3(y-1)+4)}{\sqrt{(x-2)^2+(y-1)^2}}\\
&= \lim_{(x,y)\to(2,1)} \frac{-x^2+4x-y^2+2y-5}{\sqrt{(x-2)^2+(y-1)^2}}\\
&= \lim_{(x,y)\to(2,1)} \frac{-(x-2)^2-(y-1)^2}{\sqrt{(x-2)^2+(y-1)^2}}\\
&= \lim_{(x,y)\to(2,1)} -\frac{(x-2)^2+(y-1)^2}{\sqrt{(x-2)^2+(y-1)^2}}\\
&= \lim_{(x,y)\to(2,1)} -\sqrt{(x-2)^2+(y-1)^2}\\
&= 0.
\end{aligned}$$

Por otra parte, se tiene que el vector

$$\begin{aligned}
\hat{u}_0 &= \frac{\nabla f(2,1)}{\|\nabla f(2,1)\|}\\
&= \frac{1}{\sqrt{(-3)^2+(-3)^2}}(-3,-3)\\
&= \left(-\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right)
\end{aligned}$$

es el vector que determina la dirección en la que, "parados" en el punto $(2,1)$*, obtenemos la máxima razón de cambio de la función* f*.*

Dicho en términos geométricos, si "cortamos" a la gráfica de f *por el plano perpendicular al plano* XY *que contiene a la recta que pasa por el punto* $(2,1)$*, y cuya dirección está dada por el vector* $(-1/\sqrt{2},-1/\sqrt{2})$*, entonces obtenemos la curva que en el punto* $(2,1,f(2,1))=(2,1,4)$ *tiene la recta tangente de máxima pendiente (ver figura 4.6).*

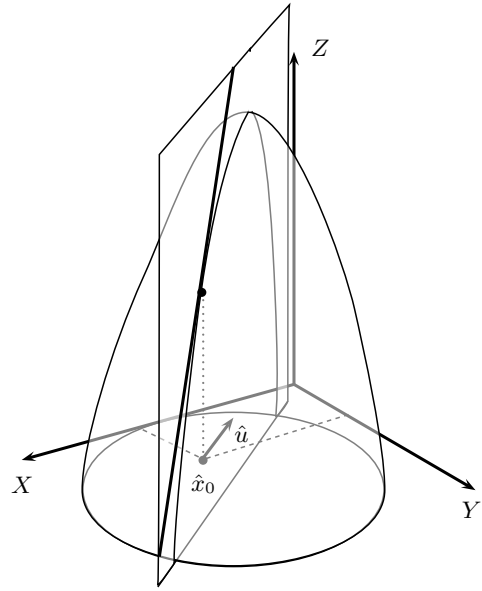


Figura 4.6: En el ejemplo 4.5, el plano perpendicular al plano XY que pasa por el punto $\hat{x}_0 = (2, 1)$, y cuya dirección está dada por el vector $\hat{u} = (-1/\sqrt{2}, -1/\sqrt{2})$ contiene a la curva y a su recta tangente, que resulta ser la de máxima pendiente que pasa por el punto $(2, 1, f(\hat{x}_0)) = (2, 1, 4)$.

4.3.2. Otras propiedades de la derivada

Una vez establecidas estas importantes propiedades del gradiente, continuaremos con el análisis de las condiciones necesarias y suficientes de la derivabilidad de una función. La siguiente propiedad que veremos es una condición necesaria, y por lo tanto nos será muy útil para determinar cuándo una función no es derivable. Esta condición es una generalización (al tipo de funciones que estamos tratando) de la propiedad que ya sabíamos para el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$, y que simplemente nos asegura que, si una función es derivable en un punto, entonces debe ser continua en ese mismo punto.

Proposición 4.12 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$. Si f es derivable en $\hat{x}_0 \in U$, entonces f es continua en $\hat{x}_0$.*

Demostración. Dado que f es derivable en $\hat{x}_0$, sabemos que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - (Df(\hat{x}_0)(\hat{x}-\hat{x}_0) + f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|} = 0$$

Como

$$f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0) = \frac{f(\hat{x}) - (Df(\hat{x}_0)(\hat{x}-\hat{x}_0) + f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|} \|\hat{x}-\hat{x}_0\|,$$

si $\hat{x} \neq \hat{x}_0$, tenemos entonces que

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} &(f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0)) \\ &= \lim_{\hat{x}\to\hat{x}_0} \left(\frac{f(\hat{x}) - (Df(\hat{x}_0)(\hat{x}-\hat{x}_0) + f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|} \|\hat{x}-\hat{x}_0\| \right) \\ &= 0.\end{aligned}$$

Por otra parte, por el problema 64 del capítulo 2 sabemos que toda función lineal es continua, de tal forma que

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} Df(\hat{x}_0)(\hat{x}-\hat{x}_0) &= \lim_{\hat{x}\to\hat{x}_0} (Df(\hat{x}_0)(\hat{x}) - Df(\hat{x}_0)(\hat{x}_0)) \\ &= 0\end{aligned}$$

Por lo tanto

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} &(f(\hat{x}) - f(\hat{x}_0)) \\ &= \lim_{\hat{x}\to\hat{x}_0} [(f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0)) + Df(\hat{x}_0)(\hat{x}-\hat{x}_0)] \\ &= \lim_{\hat{x}\to\hat{x}_0} (f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0)) + \lim_{\hat{x}\to\hat{x}_0} Df(\hat{x}_0)(\hat{x}-\hat{x}_0) \\ &= 0,\end{aligned}$$

lo que prueba que f es continua en $\hat{x}_0$. ■

Como sucede en el caso real, la condición anterior no es suficiente. Es decir, no es suficiente que una función sea continua en un punto $\hat{x}_0$ para que ésta sea derivable en dicho punto. El ejemplo que vamos a dar de esto es la "generalización" del que se suele dar para ilustrar el mismo hecho en el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$ (la función valor absoluto).

Ejemplo 4.6 *Sea* $f : \mathbb{R}^n \to \mathbb{R}$ *definida como*

$$f(\hat{x}) = \|\hat{x}\| .$$

Como sabemos, f *es continua para toda* $\hat{x} \in \mathbb{R}^n$*, en particular para* $\hat{x} = \hat{0}$*. Sin embargo, mostraremos que* f *no es derivable en* $\hat{0}$*. Para ello, primero mostraremos que en el* $\hat{0}$*, para cualquier vector* $\hat{u} \in \mathbb{R}^n$ *tal que* $\|\hat{u}\| = 1$*, la derivada direccional* $D_{\hat{u}} f(\hat{0})$ *no existe.*

En efecto, por definición sabemos que

$$\begin{aligned} D_{\hat{u}} f(\hat{0}) &= \lim_{h \to 0} \frac{f(\hat{0} + h\hat{u}) - f(\hat{0})}{h} \\ &= \lim_{h \to 0} \frac{\|h\hat{u}\| - \|\hat{0}\|}{h} \\ &= \lim_{h \to 0} \frac{|h|}{h} \end{aligned}$$

y como el lector recordará, este último límite no existe. Por lo tanto, por la proposición 4.10, f *no es derivable en el* $\hat{0}$*.*

Y ya que estamos en los contraejemplos, aprovechando la proposición anterior, mostraremos que el recíproco de la proposición 4.10 no es cierto; es decir, que la existencia de todas las derivadas direccionales de una función f en un punto $\hat{x}_0$ de su dominio no es suficiente para poder asegurar que f sea derivable en dicho punto.

Ejemplo 4.7 *Sea* $f : \mathbb{R}^2 \to \mathbb{R}$ *definida como*

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & si\ (x, y) \neq (0, 0) \\ \\ 0 & si\ (x, y) = (0, 0) \end{cases}$$

Mostraremos primero que la derivada direccional $D_{\hat{u}}f(\hat{0})$ *existe para cualquier vector* $\hat{u} = (u_1, u_2) \in \mathbb{R}^n$ *tal que* $\|\hat{u}\| = 1$.

Supongamos primero que $u_2 \neq 0$. *En este caso se tiene que*

$$\begin{aligned} D_{\hat{u}}f(\hat{0}) &= \lim_{h\to 0} \frac{f(\hat{0}+h\hat{u}) - f(\hat{0})}{h} \\ &= \lim_{h\to 0} \frac{1}{h}\left(\frac{(hu_1)^2(hu_2)}{(hu_1)^4 + (hu_2)^2}\right) \\ &= \lim_{h\to 0} \frac{1}{h}\left(\frac{h^3u_1^2u_2}{h^2\left(h^2u_1^4 + u_2^2\right)}\right) \\ &= \lim_{h\to 0} \frac{u_1^2u_2}{h^2u_1^4 + u_2^2} \\ &= \frac{u_1^2}{u_2} \end{aligned}$$

y si $u_2 = 0$, *es decir que* $\hat{u} = (1, 0) = \hat{e}_1$, *entonces*

$$\begin{aligned} D_{\hat{u}}f(\hat{0}) &= \lim_{h\to 0} \frac{f(\hat{0}+h\hat{u}) - f(\hat{0})}{h} \\ &= \lim_{h\to 0} \frac{f(h, 0) - f(0, 0)}{h} \\ &= \lim_{h\to 0} \frac{0-0}{h} = 0. \end{aligned}$$

Esto prueba que $D_{\hat{u}}f(\hat{0})$ *existe para cualquier vector* $\hat{u} \in \mathbb{R}^n$ *tal que* $\|\hat{u}\| = 1$.

Por otra parte, si evaluamos a f *en puntos de la forma* (t, t^2), *con* $t \neq 0$, *se tiene que*

$$\begin{aligned} f(t, t^2) &= \frac{(t)^2t^2}{(t)^4 + (t^2)^2} \\ &= \frac{t^4}{t^4 + t^4} \\ &= \frac{1}{2}, \end{aligned}$$

de tal forma que, como $(t, t^2) \to (0,0)$ *si* $t \to 0$, *concluimos que* f *no es continua en el* $\hat{0}$ *y por la proposición 4.12, que* f *no es derivable en el* $\hat{0}$.

Si bien es cierto que la existencia de las derivadas direccionales de una función f en un punto $\hat{x}_0$ no es suficiente para garantizar su derivabilidad en dicho punto, un análisis más de cerca del ejemplo anterior nos puede arrojar luz sobre cuáles hipótesis adicionales habría que considerar para poder asegurar la derivabilidad de f en $\hat{x}_0$.

Lo primero que haremos será mostrar que la función del ejemplo anterior sí es derivable en cualquier punto $\hat{x} = (x, y) \neq \hat{0}$. Para ello, como ya sabemos, será necesario calcular las derivadas parciales $\partial f/\partial x$ y $\partial f/\partial y$ en $\hat{x}$ y mostrar que la función lineal representada por la matriz

$$\left[\begin{array}{cc} \frac{\partial f}{\partial x}(x,y) & \frac{\partial f}{\partial y}(x,y) \end{array} \right]$$

satisface la condición establecida en la definición 4.3.

Con base en varios de los incisos de la proposición 4.6 y en los cálculos realizados en el ejemplo anterior, tenemos que

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} \frac{2xy(y^2-x^4)}{(x^4+y^2)^2} & \text{si } (x,y) \neq (0,0) \\ \\ 0 & \text{si } (x,y) = (0,0) \end{cases} \tag{4.24}$$

y

$$\frac{\partial f}{\partial y}(x,y) = \begin{cases} \frac{x^2(x^4-y^2)}{(x^4+y^2)^2} & \text{si } (x,y) \neq (0,0) \\ \\ 0 & \text{si } (x,y) = (0,0) \end{cases} \tag{4.25}$$

Si ahora tomamos $\hat{x}_0 = (x_0, y_0) \neq (0,0)$, para cualquier otra $\hat{x} = (x, y) \in \mathbb{R}^2$ se tiene que

$$\begin{aligned} &|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))| \\ &= \left| f(x,y) - f(x_0,y_0) - \frac{\partial f}{\partial x}(\hat{x}_0)(x - x_0) - \frac{\partial f}{\partial y}(\hat{x}_0)(y - y_0) \right| \end{aligned}$$

$$= |f(x,y) - f(x_0,y) + f(x_0,y) - f(x_0,y_0) \\ -\frac{\partial f}{\partial x}(\hat{x}_0)(x-x_0) - \frac{\partial f}{\partial y}(\hat{x}_0)(y-y_0)\Big|.$$

Dado que las derivadas parciales $\partial f/\partial x$ y $\partial f/\partial y$ existen para toda $\hat{x} \in \mathbb{R}^2$, por la proposición 4.8 sabemos que

$$f(x,y) - f(x_0,y) = (x-x_0)\,\frac{\partial f}{\partial x}(\xi_x, y)$$

y

$$f(x_0,y) - f(x_0,y_0) = (y-y_0)\,\frac{\partial f}{\partial y}(x_0, \xi_y),$$

y por tanto

$$\begin{aligned}
&|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x}-\hat{x}_0) + f(\hat{x}_0))| \\
&= |f(x,y) - f(x_0,y) + f(x_0,y) - f(x_0,y_0) \\
&\qquad -\frac{\partial f}{\partial x}(\hat{x}_0)(x-x_0) - \frac{\partial f}{\partial y}(\hat{x}_0)(y-y_0)\Big| \\
&= \Big|(x-x_0)\,\frac{\partial f}{\partial x}(\xi_x, y) + (y-y_0)\,\frac{\partial f}{\partial y}(x_0,\xi_y) \\
&\qquad -\frac{\partial f}{\partial x}(\hat{x}_0)(x-x_0) - \frac{\partial f}{\partial y}(\hat{x}_0)(y-y_0)\Big| \\
&= \left|\left(\frac{\partial f}{\partial x}(\xi_x,y) - \frac{\partial f}{\partial x}(\hat{x}_0)\right)(x-x_0) + \left(\frac{\partial f}{\partial y}(x_0,\xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0)\right)(y-y_0)\right| \\
&\leq \left|\frac{\partial f}{\partial x}(\xi_x,y) - \frac{\partial f}{\partial x}(\hat{x}_0)\right| |x-x_0| + \left|\frac{\partial f}{\partial y}(x_0,\xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0)\right| |y-y_0| \\
&\leq \left|\frac{\partial f}{\partial x}(\xi_x,y) - \frac{\partial f}{\partial x}(\hat{x}_0)\right| \|\hat{x}-\hat{x}_0\| + \left|\frac{\partial f}{\partial y}(x_0,\xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0)\right| \|\hat{x}-\hat{x}_0\| \\
&= \left(\left|\frac{\partial f}{\partial x}(\xi_x,y) - \frac{\partial f}{\partial x}(\hat{x}_0)\right| + \left|\frac{\partial f}{\partial y}(x_0,\xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0)\right|\right) \|\hat{x}-\hat{x}_0\|.
\end{aligned}$$

De aquí obtenemos que

$$\begin{aligned} 0 &\leq \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} \\ &\leq \left|\frac{\partial f}{\partial x}(\xi_x, y) - \frac{\partial f}{\partial x}(\hat{x}_0)\right| + \left|\frac{\partial f}{\partial y}(x_0, \xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0)\right|. \end{aligned}$$

Si ahora notamos que, de acuerdo con las expresiones 4.24 y 4.25 de las derivadas parciales de f, éstas son continuas en cualquier punto $\hat{x}_0 \neq \hat{0}$, y que los puntos (ξ_x, y) y (x_0, ξ_y) tienden al punto $(x_0, y_0) = \hat{x}_0$ si el punto $\hat{x} = (x, y)$ tiende al punto $(x_0, y_0) = \hat{x}_0$, tenemos que

$$\begin{aligned} 0 &\leq \lim_{\hat{x}\to\hat{x}_0} \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} \\ &\leq \lim_{\hat{x}\to\hat{x}_0} \left(\left|\frac{\partial f}{\partial x}(\xi_x, y) - \frac{\partial f}{\partial x}(\hat{x}_0)\right| + \left|\frac{\partial f}{\partial y}(x_0, \xi_y) - \frac{\partial f}{\partial y}(\hat{x}_0)\right|\right) \\ &= 0 + 0 \\ &= 0, \end{aligned}$$

de donde concluimos que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} = 0,$$

o equivalentemente, que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} = 0,$$

es decir, que f es derivable en $\hat{x}_0$.

Ojalá y el lector esté de acuerdo en que el arduo trabajo que nos tomamos para probar que la función del ejemplo 4.7 es derivable para cualquier $\hat{x}_0 \in \mathbb{R}^2$ distinto del $\hat{0}$ será redituable. En efecto, nótese que más allá de la forma explícita de f, lo importante en la argumentación anterior fueron dos cosas:

1. que las derivadas parciales de f existieron para todos los puntos de $\mathbb{R}^2$, y

2. que fueron continuas en $\hat{x}_0$.

Con respecto al punto 1, el lector estará de acuerdo (dado el caracter "local" de la derivada), en que en el desarrollo anterior hubiera bastado con que las derivadas parciales existieran en una vecindad del punto $\hat{x}_0$. Pues bien, lo que probaremos a continuación es que estas dos condiciones (con la modificación propuesta a la primera) son suficientes para poder asegurar que una función es derivable en un punto $\hat{x}_0$, lo que dejaremos formulado en la siguiente

Proposición 4.13 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Si $\frac{\partial f}{\partial x_i}(\hat{x})$ existe para cada $\hat{x} \in B_r(\hat{x}_0)$ y $\frac{\partial f}{\partial x_i}$ es continua en $\hat{x}_0$ (para cada $i \in \{1, \ldots, n\}$), entonces f es derivable en $\hat{x}_0$.*

Demostración. Sea $\hat{x} \in B_r(\hat{x}_0)$, $\hat{x} \neq \hat{x}_0$. Si $\hat{x}_0 = \left(x_1^{(0)}, \ldots, x_n^{(0)}\right)$ y $\hat{x} = (x_1, \ldots, x_n)$, hacemos

$$\hat{x}_i = \left(x_1, \ldots, x_i, x_{i+1}^{(0)}, \ldots, x_n^{(0)}\right)$$

para cada $i \in \{1, \ldots, n-1\}$ y $\hat{x}_n = \hat{x}$. Por el inciso (a) del problema 11 del capítulo 1 sabemos que $\hat{x}_i \in B_r(\hat{x}_0)$ para cada $i \in \{1, \ldots, n-1\}$ y como $B_r(\hat{x}_0)$ es un conjunto convexo, entonces los segmentos $[\hat{x}_{i-1}, \hat{x}_i] \subset B_r(\hat{x}_0)$ para cada $i \in \{1, \ldots, n\}$.

Por otra parte, nótese que

$$\begin{aligned} f(\hat{x}) - f(\hat{x}_0) &= (f(\hat{x}) - f(\hat{x}_{n-1})) + (f(\hat{x}_{n-1}) - f(\hat{x}_{n-2})) + \cdots \\ &\qquad + (f(\hat{x}_2) - f(\hat{x}_1)) + (f(\hat{x}_1) - f(\hat{x}_0)) \\ &= \sum_{i=1}^{n} (f(\hat{x}_i) - f(\hat{x}_{i-1})). \end{aligned}$$

Ahora, dado que

$$\hat{x}_i - \hat{x}_{i-1} = \left(x_i - x_i^{(0)}\right) \hat{e}_i,$$

por la proposición 4.8 sabemos que existe $\hat{\xi}_i \in [\hat{x}_{i-1}, \hat{x}_i]$ tal que

$$f(\hat{x}_i) - f(\hat{x}_{i-1}) = \left(x_i - x_i^{(0)}\right) \frac{\partial f}{\partial x_i}\left(\hat{\xi}_i\right)$$

para cada $i \in \{1, \ldots, n\}$, de modo que

$$\begin{aligned} f(\hat{x}) - f(\hat{x}_0) &= \sum_{i=1}^{n} (f(\hat{x}_i) - f(\hat{x}_{i-1})) \\ &= \sum_{i=1}^{n} \left(x_i - x_i^{(0)}\right) \frac{\partial f}{\partial x_i}\left(\hat{\xi}_i\right) \\ &= \left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) \cdot (\hat{x} - \hat{x}_0). \end{aligned}$$

Por lo tanto

$$\begin{aligned} &|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))| \\ &= |f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)| \end{aligned}$$

$$\begin{aligned} &= \left| \left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) \cdot (\hat{x} - \hat{x}_0) - \nabla f(\hat{x}_0) \cdot (\hat{x} - \hat{x}_0) \right| \\ &= \left| \left(\left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) - \nabla f(\hat{x}_0) \right) \cdot (\hat{x} - \hat{x}_0) \right| \\ &\leq \left\| \left(\left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) - \nabla f(\hat{x}_0) \right) \right\| \|(\hat{x} - \hat{x}_0)\|, \end{aligned}$$

de modo que

$$\begin{aligned} 0 &\leq \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} \\ &\leq \left\| \left(\left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right)\right) - \nabla f(\hat{x}_0) \right) \right\|. \end{aligned}$$

Si ahora recordamos que por el inciso (b) del problema 11 del capítulo 1 se tiene que

$$\left\|\hat{\xi}_i - \hat{x}_0\right\| \leq \|\hat{x} - \hat{x}_0\|,$$

entonces $\hat{\xi}_i \to \hat{x}_0$ si $\hat{x} \to \hat{x}_0$, de modo que, como $\frac{\partial f}{\partial x_i}$ es continua en $\hat{x}_0$ (para cada $i \in \{1, \ldots, n\}$), tenemos que

$$\begin{aligned}
0 &\leq \lim_{\hat{x}\to\hat{x}_0} \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} \\
&\leq \lim_{\hat{x}\to\hat{x}_0} \left\| \left(\left(\frac{\partial f}{\partial x_1}\left(\hat{\xi}_1\right), \frac{\partial f}{\partial x_2}\left(\hat{\xi}_2\right), \ldots, \frac{\partial f}{\partial x_2}\left(\hat{\xi}_n\right) \right) - \nabla f(\hat{x}_0) \right) \right\| \\
&= \|\nabla f(\hat{x}_0) - \nabla f(\hat{x}_0)\| \\
&= 0.
\end{aligned}$$

Es decir,

$$\lim_{\hat{x}\to\hat{x}_0} \frac{|f(\hat{x}) - (Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0))|}{\|\hat{x} - \hat{x}_0\|} = 0,$$

lo cual prueba que f es derivable en $\hat{x}_0$. ■

Es importante hacer notar que la condición de la proposición anterior sólo es una condición suficiente para la derivabilidad de f en $\hat{x}_0$; es decir, si bien es cierto que la existencia de la derivada de f en $\hat{x}_0$ garantiza la existencia de todas sus derivadas direccionales en este punto (incluyendo sus derivadas parciales), este hecho ni siquiera garantiza que existan todas las derivadas parciales de f en todos los puntos de alguna vecindad de $\hat{x}_0$.

Esta última afirmación queda ilustrada en el siguiente

Ejemplo 4.8 *Sea* $f : \mathbb{R}^2 \to \mathbb{R}$ *definida como*

$$f(x, y) = |xy|.$$

Nótese que

$$\begin{aligned}\frac{\partial f}{\partial x}(0,0) &= \lim_{h\to 0}\frac{f(h,0)-f(0,0)}{h}\\ &= \lim_{h\to 0}\frac{|h\cdot 0|-|0\cdot 0|}{h}\\ &= 0\end{aligned}$$

y que

$$\begin{aligned}\frac{\partial f}{\partial y}(0,0) &= \lim_{h\to 0}\frac{f(0,h)-f(0,0)}{h}\\ &= \lim_{h\to 0}\frac{|0h|-|0\cdot 0|}{h}\\ &= 0.\end{aligned}$$

Por lo tanto, tomando a $Df(0,0)$ *como la constante* 0*, tenemos que*

$$\begin{aligned}0 &\leq \lim_{(x,y)\to(0,0)}\frac{f(x,y)-(Df(0,0)((x,y)-(0,0))+f(0,0))}{\|(x,y)-(0,0)\|}\\ &= \lim_{(x,y)\to(0,0)}\frac{|xy|}{\|(x,y)\|}\\ &\leq \lim_{(x,y)\to(0,0)}\frac{\|(x,y)\|^2}{\|(x,y)\|}\\ &= \lim_{(x,y)\to(0,0)}\|(x,y)\|\\ &= 0,\end{aligned}$$

de modo que

$$\lim_{(x,y)\to(0,0)}\frac{f(x,y)-(Df(0,0)((x,y)-(0,0))+f(0,0))}{\|(x,y)-(0,0)\|}=0,$$

lo que prueba que f *es derivable en el* $(0,0)$*.*

Por otra parte, nótese que tomando cualquier punto de la forma $(x_0, 0)$*, con* $x_0 \neq 0$*, de acuerdo con la definición de derivada parcial, se tiene que*

$$\begin{aligned}\frac{\partial f}{\partial y}(x_0, 0) &= \lim_{h\to 0} \frac{f(x_0, h) - f(x_0, 0)}{h} \\ &= \lim_{h\to 0} \frac{|x_0 h| - 0}{h} \\ &= |x_0| \lim_{h\to 0} \frac{|h|}{h}.\end{aligned}$$

Como este último límite no existe, entonces $\frac{\partial f}{\partial y}(x_0, 0)$ *no existe.*

Análogamente, para cualquier punto de la forma $(0, y_0)$*, con* $y_0 \neq 0$*, se tiene que*

$$\begin{aligned}\frac{\partial f}{\partial x}(0, y_0) &= \lim_{h\to 0} \frac{f(h, y_0) - f(0, y_0)}{h} \\ &= \lim_{h\to 0} \frac{|h y_0| - 0}{h} \\ &= |y_0| \lim_{h\to 0} \frac{|h|}{h},\end{aligned}$$

de modo que $\frac{\partial f}{\partial x}(0, y_0)$ *tampoco existe.*

Por tanto, en cualquier vecindad del $(0, 0)$ *se tiene que existen puntos para los cuales* $\frac{\partial f}{\partial x}$ *y* $\frac{\partial f}{\partial y}$ *no existen, que es lo que deseábamos mostrar (ver figura 4.7).*

Aun cuando el ejemplo anterior muestra sin lugar a dudas que la proposición 4.13 sólo nos proporciona una condición suficiente para la derivabilidad de una función en un punto, nos tomaremos el trabajo de dar otro ejemplo en el que las derivadas parciales, a diferencia del ejemplo anterior, sí existen

en una vecindad del punto en cuestión, pero no son continuas en el punto en el que la función sí es derivable.

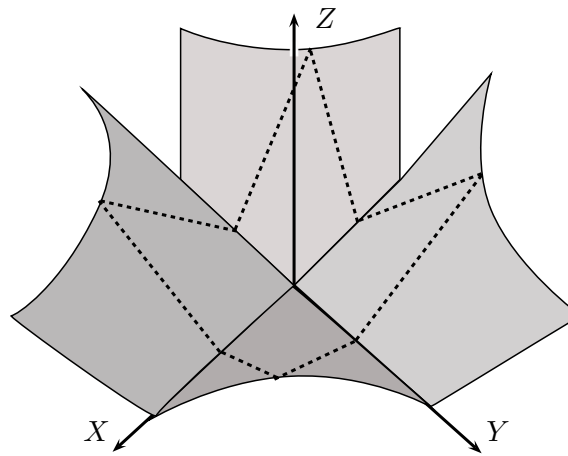


Figura 4.7: Gráfica de la función $f(x,y) = |xy|$ (del ejemplo 4.8) cuyas derivadas parciales no existen en los puntos de la forma $(x,0)$, con $x \neq 0$ y $(0,y)$, con $y \neq 0$.

Ejemplo 4.9 *Sea* $f : \mathbb{R}^2 \to \mathbb{R}$ *definida como*

$$f(x,y) = \begin{cases} (x^2+y^2)\operatorname{sen}\left(\frac{1}{\sqrt{x^2+y^2}}\right) & si\ (x,y) \neq (0,0) \\ \\ 0 & si\ (x,y) = (0,0) \end{cases}$$

En este caso tenemos que

$$\begin{aligned} \frac{\partial f}{\partial x}(0,0) &= \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} \\ &= \lim_{h\to 0} h \operatorname{sen}\left(\frac{1}{|h|}\right) \\ &= 0 \end{aligned}$$

y

$$\begin{aligned}\frac{\partial f}{\partial y}(0,0) &= \lim_{h\to 0}\frac{f(0,h)-f(0,0)}{h}\\ &= \lim_{h\to 0} h\,\operatorname{sen}\left(\frac{1}{|h|}\right)\\ &= 0.\end{aligned}$$

Si $(x,y)\neq(0,0)$*, con base en la proposición 4.6, obtenemos que*

$$\frac{\partial f}{\partial x}(x,y) = 2x\,\operatorname{sen}\left(\frac{1}{\sqrt{x^2+y^2}}\right) - \frac{x}{\sqrt{x^2+y^2}}\cos\left(\frac{1}{\sqrt{x^2+y^2}}\right)$$

y

$$\frac{\partial f}{\partial y}(x,y) = 2y\,\operatorname{sen}\left(\frac{1}{\sqrt{x^2+y^2}}\right) - \frac{y}{\sqrt{x^2+y^2}}\cos\left(\frac{1}{\sqrt{x^2+y^2}}\right).$$

Si ahora tomamos puntos de la forma $(t,0)$*, con* $t>0$*, se tiene que*

$$\frac{\partial f}{\partial x}(t,0) = 2t\,\operatorname{sen}\left(\frac{1}{t}\right) - \cos\left(\frac{1}{t}\right)$$

y como

$$\lim_{t\to 0}\left(2t\,\operatorname{sen}\left(\frac{1}{t}\right) - \cos\left(\frac{1}{t}\right)\right)$$

no existe, concluimos que

$$\lim_{(x,y)\to(0,0)}\frac{\partial f}{\partial x}(x,y)$$

tampoco existe, de modo que $\frac{\partial f}{\partial x}$ *no es continua en el* $(0,0)$*. Análogamente, evaluando en puntos de la forma* $(0,t)$*, con* $t>0$*, concluimos que* $\frac{\partial f}{\partial y}$ *no es continua en el* $(0,0)$*.*

Por otra parte, se tiene que

$$\begin{aligned}\lim_{(x,y)\to(0,0)} & \frac{f(x,y)-(Df(0,0)((x,y)-(0,0))+f(0,0))}{\|(x,y)-(0,0)\|}\\ &= \lim_{(x,y)\to(0,0)} \frac{(x^2+y^2)\operatorname{sen}\left(\frac{1}{\sqrt{x^2+y^2}}\right)}{\|(x,y)\|}\\ &= \lim_{(x,y)\to(0,0)} \|(x,y)\|\operatorname{sen}\left(\frac{1}{\sqrt{x^2+y^2}}\right)\\ &= 0,\end{aligned}$$

lo que demuestra que f sí es derivable en el $(0,0)$.

Con el ejemplo anterior concluimos el análisis de las condiciones necesarias y suficientes relacionadas con la derivabilidad de una función. Lo siguiente que haremos será mostrar la relación entre este concepto y la aritmética de las funciones, la cual dejaremos plasmada en la siguiente

Proposición 4.14 *Sean $f, g : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Si f y g son derivables en $\hat{x}_0$, se satisfacen las siguientes afirmaciones:*

1. *$f+g$ es derivable en $\hat{x}_0$ y además*

$$D(f+g)(\hat{x}_0) = Df(\hat{x}_0) + Dg(\hat{x}_0)$$

2. *si $\alpha \in \mathbb{R}$, entonces αf es derivable en $\hat{x}_0$ y además*

$$D(\alpha f)(\hat{x}_0) = \alpha Df(\hat{x}_0)$$

3. *fg es derivable en $\hat{x}_0$ y además*

$$D(fg)(\hat{x}_0) = f(\hat{x}_0)Dg(\hat{x}_0) + g(\hat{x}_0)Df(\hat{x}_0)$$

4. *si $g(\hat{x}_0) \neq 0$, entonces f/g es derivable en $\hat{x}_0$ y además*

$$D(f/g)(\hat{x}_0) = \frac{1}{g^2(\hat{x}_0)}\left(g(\hat{x}_0)Df(\hat{x}_0) - f(\hat{x}_0)Dg(\hat{x}_0)\right)$$

Demostración. A manera de ejemplo, probaremos el inciso 3 y el resto quedarán como problemas para el lector. Una vez dicho lo anterior, nótese que

$$\begin{aligned}
&(fg)(\hat{x}) - (fg)(\hat{x}_0) - \left[f(\hat{x}_0)Dg(\hat{x}_0) + g(\hat{x}_0)Df(\hat{x}_0)\right](\hat{x} - \hat{x}_0) \\
&= f(\hat{x})\left[g(\hat{x}) - g(\hat{x}_0)\right] + g(\hat{x}_0)\left[f(\hat{x}) - f(\hat{x}_0)\right] \\
&\qquad\qquad - \left[f(\hat{x}_0)Dg(\hat{x}_0) + g(\hat{x}_0)Df(\hat{x}_0)\right](\hat{x} - \hat{x}_0) \\
&= f(\hat{x})\left[g(\hat{x}) - g(\hat{x}_0) - Dg(\hat{x}_0)(\hat{x} - \hat{x}_0)\right] + \left[f(\hat{x}) - f(\hat{x}_0)\right]Dg(\hat{x}_0)(\hat{x} - \hat{x}_0) \\
&\qquad\qquad + g(\hat{x}_0)\left[f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)\right],
\end{aligned}$$

de tal forma que

$$\begin{aligned}
&\frac{(fg)(\hat{x}) - (fg)(\hat{x}_0) - \left(f(\hat{x}_0)Dg(\hat{x}_0) + g(\hat{x}_0)Df(\hat{x}_0)\right)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \\
&= f(\hat{x})\frac{g(\hat{x}) - g(\hat{x}_0) - Dg(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} + \left[f(\hat{x}) - f(\hat{x}_0)\right]\frac{Dg(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \\
&+ g(\hat{x}_0)\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|}.
\end{aligned}$$

Para el primer y tercer sumando que aparecen en el lado derecho de esta última identidad, dado que f y g son derivables en $\hat{x}_0$ (y por lo tanto continuas en ese mismo punto), podemos concluir que

$$\begin{aligned}
&\lim_{\hat{x}\to\hat{x}_0} f(\hat{x})\frac{g(\hat{x}) - g(\hat{x}_0) - Dg(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \\
&\qquad = \left(\lim_{\hat{x}\to\hat{x}_0} f(\hat{x})\right)\left(\lim_{\hat{x}\to\hat{x}_0} \frac{g(\hat{x}) - g(\hat{x}_0) - Dg(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|}\right) \\
&\qquad = f(\hat{x}_0)0 \\
&\qquad = 0
\end{aligned}$$

y

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} g(\hat{x}_0)\frac{f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)\,(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}&\\ = g(\hat{x}_0)\lim_{\hat{x}\to\hat{x}_0}\frac{f(\hat{x})-f(\hat{x}_0)-Df(\hat{x}_0)\,(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}&\\ = g(\hat{x}_0)0&\\ = 0.&\end{aligned}$$

Por otra parte, por el inciso (c) del problema 11 del capítulo 2, sabemos que existe $M \geq 0$ tal que

$$|Df(\hat{x}_0)\,(\hat{x}-\hat{x}_0)| \leq M\,\|\hat{x}-\hat{x}_0\|$$

para toda $\hat{x} \in \mathbb{R}^n$, y como f es continua en $\hat{x}_0$, concluimos que

$$\begin{aligned}0 &\leq \lim_{\hat{x}\to\hat{x}_0}\left|[f(\hat{x})-f(\hat{x}_0)]\,\frac{Dg(\hat{x}_0)\,(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\right|\\ &\leq M\lim_{\hat{x}\to\hat{x}_0}|f(\hat{x})-f(\hat{x}_0)|\\ &= M0\\ &= 0\end{aligned}$$

y por lo tanto que

$$\lim_{\hat{x}\to\hat{x}_0}[f(\hat{x})-f(\hat{x}_0)]\,\frac{Dg(\hat{x}_0)\,(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}=0.$$

Sumando estos límites, concluimos que

$$\lim_{\hat{x}\to\hat{x}_0}\frac{(fg)(\hat{x})-(fg)(\hat{x}_0)-(f(\hat{x}_0)Dg(\hat{x}_0)+g(\hat{x}_0)Df(\hat{x}_0))\,(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}=0,$$

de modo que la función fg es derivable en $\hat{x}_0$ y además que

$$D(fg)(\hat{x}_0)=f(\hat{x}_0)Dg(\hat{x}_0)+g(\hat{x}_0)Df(\hat{x}_0).$$

■

Para terminar con las propiedades de la derivada relacionadas con las operaciones entre funciones, formularemos (en dos proposiciones separadas) aquellas que nos hablan de la composición de funciones.

Serán dos proposiciones diferentes puesto que existen dos formas de componer a una función de $\mathbb{R}^n$ en $\mathbb{R}$ con funciones para las cuales también tenemos una definición de derivada: por la izquierda, con una función de $\mathbb{R}$ en $\mathbb{R}$, o por la derecha, con una función de $\mathbb{R}$ en $\mathbb{R}^n$. Ambas proposiciones serán un caso particular de un resultado más general que veremos más adelante (y que se le conoce con el nombre de "regla de la cadena").

Proposición 4.15 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $(a,b) \subset \mathbb{R}$ tal que $f(U) \subset (a,b)$ y $g : (a,b) \subset \mathbb{R} \to \mathbb{R}$. Si f es derivable en $\hat{x}_0$ y g es derivable en $f(\hat{x}_0)$, entonces $g \circ f$ es derivable en $\hat{x}_0$ y además*

$$D(g \circ f)(\hat{x}_0) = g'(f(\hat{x}_0))Df(\hat{x}_0)$$

Demostración. Recordemos que, de acuerdo con la definición 4.3, lo que tenemos que demostrar es que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{(g \circ f)(\hat{x}) - (g \circ f)(\hat{x}_0) - g'(f(\hat{x}_0))Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} = 0,$$

razón por la cual empezaremos por buscar una expresión "adecuadamente equivalente" al numerador del cociente anterior.

Para ello, definamos

$$k_{\hat{x}} = f(\hat{x}) - f(\hat{x}_0)$$

y observemos que, como f es derivable en $\hat{x}_0$, entonces f es continua en $\hat{x}_0$ (proposición 4.12), de donde se tiene que

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}_0} k_{\hat{x}} &= \lim_{\hat{x}\to\hat{x}_0} (f(\hat{x}) - f(\hat{x}_0)) \\ &= 0.\end{aligned}$$

En el caso en que $k_{\hat{x}} \neq 0$, tenemos que

$$\begin{aligned}
&(g \circ f)(\hat{x}) - (g \circ f)(\hat{x}_0) \\
&= g(f(\hat{x})) - g(f(\hat{x}_0)) \\
&= \frac{g(f(\hat{x}_0) + k_{\hat{x}}) - g(f(\hat{x}_0))}{k_{\hat{x}}} k_{\hat{x}} \\
&= \frac{g(f(\hat{x}_0) + k_{\hat{x}}) - g(f(\hat{x}_0))}{k_{\hat{x}}} (f(\hat{x}) - f(\hat{x}_0)) \\
&= \frac{g(f(\hat{x}_0) + k_{\hat{x}}) - g(f(\hat{x}_0))}{k_{\hat{x}}} (f(\hat{x}) - f(\hat{x}_0) \\
&\qquad\qquad - Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + Df(\hat{x}_0)(\hat{x} - \hat{x}_0)) \\
&= \frac{g(f(\hat{x}_0) + k_{\hat{x}}) - g(f(\hat{x}_0))}{k_{\hat{x}}} (f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)) \\
&+ \frac{g(f(\hat{x}_0) + k_{\hat{x}}) - g(f(\hat{x}_0))}{k_{\hat{x}}} Df(\hat{x}_0)(\hat{x} - \hat{x}_0),
\end{aligned}$$

de tal manera que

$$\begin{aligned}
&(g \circ f)(\hat{x}) - (g \circ f)(\hat{x}_0) - g'(f(\hat{x}_0)) Df(\hat{x}_0)(\hat{x} - \hat{x}_0) \\
&= \frac{g(f(\hat{x}_0) + k_{\hat{x}}) - g(f(\hat{x}_0))}{k_{\hat{x}}} (f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)) \\
&+ \left(\frac{g(f(\hat{x}_0) + k_{\hat{x}}) - g(f(\hat{x}_0))}{k_{\hat{x}}} - g'(f(\hat{x}_0)) \right) Df(\hat{x}_0)(\hat{x} - \hat{x}_0).
\end{aligned}$$

Por lo tanto

$$\begin{aligned}
&\frac{(g \circ f)(\hat{x}) - (g \circ f)(\hat{x}_0) - g'(f(\hat{x}_0)) Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \\
&= \frac{g(f(\hat{x}_0) + k_{\hat{x}}) - g(f(\hat{x}_0))}{k_{\hat{x}}} \left(\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \right) \\
&+ \left(\frac{g(f(\hat{x}_0) + k_{\hat{x}}) - g(f(\hat{x}_0))}{k_{\hat{x}}} - g'(f(\hat{x}_0)) \right) Df(\hat{x}_0) \left(\frac{\hat{x} - \hat{x}_0}{\|\hat{x} - \hat{x}_0\|} \right).
\end{aligned} \tag{4.26}$$

Como seguramente el lector ya habrá notado, cada uno de los sumandos del lado derecho de esta última identidad tienden a 0 cuando $\hat{x} \to \hat{x}_0$.

En efecto, los factores del primer sumando tienen límite; el primero tiende a $g'(f(\hat{x}_0))$ (puesto que g es derivable en $f(\hat{x}_0)$) y el segundo tiende a 0 (puesto que f es derivable en $\hat{x}_0$).

En cuanto al segundo sumando, su primer factor tiende a 0 (nuevamente porque g es derivable en $f(\hat{x}_0)$) mientras que para el segundo factor, por el inciso (c) del problema 11 del capítulo 2, sabemos que existe $M \geq 0$ tal que

$$|Df(\hat{x}_0)\,(\hat{x} - \hat{x}_0)| \leq M\,\|\hat{x} - \hat{x}_0\|\,,$$

de tal modo que

$$\left| Df(\hat{x}_0)\left(\frac{\hat{x} - \hat{x}_0}{\|\hat{x} - \hat{x}_0\|}\right)\right| \leq M$$

y por lo tanto todo el sumando tiende a 0 cuando $\hat{x} \to \hat{x}_0$.

Toda la argumentación anterior es correcta bajo el supuesto de que $k_{\hat{x}} = f(\hat{x}) - f(\hat{x}_0) \neq 0$; lo que ahora mostraremos es que para considerar el caso en que $k_{\hat{x}} = 0$ será necesario introducir una función auxiliar[1] $\varphi : U \subset \mathbb{R}^n \to \mathbb{R}$ definida de la siguiente forma

$$\varphi(\hat{x}) = \begin{cases} \frac{g(f(\hat{x})) - g(f(\hat{x}_0))}{f(\hat{x}) - f(\hat{x}_0)} & \text{si } f(\hat{x}) - f(\hat{x}_0) \neq 0 \\ \\ g'(f(\hat{x}_0)) & \text{si } f(\hat{x}) - f(\hat{x}_0) = 0 \end{cases}$$

Lo primero que haremos notar es que la igualdad 4.26 se escribe en términos de la función φ como

$$\begin{aligned} &\frac{(g \circ f)(\hat{x}) - (g \circ f)(\hat{x}_0) - g'(f(\hat{x}_0))Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \\ &= \varphi(\hat{x})\left(\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|}\right) \\ &\qquad + \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) Df(\hat{x}_0)\left(\frac{\hat{x} - \hat{x}_0}{\|\hat{x} - \hat{x}_0\|}\right) \end{aligned}$$

[1] Esto es lo que siempre se hace en las pruebas de las diferentes variantes de la regla de la cadena, como seguramente el lector ya habrá notado a estas alturas.

y que ahora esta identidad se cumple para toda $\hat{x} \in U$, $\hat{x} \neq \hat{x}_0$. Si ahora recordamos que f es continua en $\hat{x}_0$, dado que g es derivable en $f(\hat{x}_0)$, aplicando el resultado del problema 48 del capítulo 2, sabemos que φ es continua en $\hat{x}_0$ y por tanto

$$\begin{aligned}
&\lim_{\hat{x}\to\hat{x}_0} \frac{(g\circ f)(\hat{x}) - (g\circ f)(\hat{x}_0) - g'(f(\hat{x}_0))Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|} \\
&= \lim_{\hat{x}\to\hat{x}_0} \left[\varphi(\hat{x}) \left(\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\right)\right. \\
&\qquad\qquad \left. + \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) Df(\hat{x}_0) \left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right)\right] \\
&= \lim_{\hat{x}\to\hat{x}_0} \varphi(\hat{x}) \left(\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)(\hat{x}-\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|}\right) \\
&\qquad\qquad + \lim_{\hat{x}\to\hat{x}_0} \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) Df(\hat{x}_0) \left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right) \\
&= \varphi(\hat{x}_0)0 + \lim_{\hat{x}\to\hat{x}_0} \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) Df(\hat{x}_0) \left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right) \\
&= \lim_{\hat{x}\to\hat{x}_0} \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) Df(\hat{x}_0) \left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right) \\
&= 0.
\end{aligned}$$

La última identidad se satisface dado que

$$\begin{aligned}
\lim_{\hat{x}\to\hat{x}_0} \left(\varphi(\hat{x}) - g'(f(\hat{x}_0))\right) &= \lim_{\hat{x}\to\hat{x}_0} \varphi(\hat{x}) - g'(f(\hat{x}_0)) \\
&= \varphi(\hat{x}_0) - g'(f(\hat{x}_0)) \\
&= g'(f(\hat{x}_0)) - g'(f(\hat{x}_0)) \\
&= 0
\end{aligned}$$

y que, como mencionamos párrafos arriba, existe $M \geq 0$ tal que

$$\left| Df(\hat{x}_0) \left(\frac{\hat{x}-\hat{x}_0}{\|\hat{x}-\hat{x}_0\|}\right)\right| \leq M$$

para toda $\hat{x} \in \mathbb{R}^n$, $\hat{x} \neq \hat{x}_0$. ■

Como afirmamos anteriormente, también podemos componer (por la derecha) a una función f de $\mathbb{R}^n$ en $\mathbb{R}$, con una función γ de $\mathbb{R}$ en $\mathbb{R}^n$. Para esta composición también podemos establecer una regla de la cadena: bajo la hipótesis de que ambas funciones son derivables (en los puntos adecuados), entonces dicha composición también es derivable, y además nos da una forma de cómo calcular esta derivada.

En la formulación (y prueba) de este resultado representaremos a la derivada de f a través de su vector gradiente.

Proposición 4.16 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $\gamma : (a,b) \subset \mathbb{R} \to \mathbb{R}^n$ tal que $\gamma(a,b) \subset U$, y $t_0 \in (a,b)$ tal que $\gamma(t_0) = \hat{x}_0$. Si f es derivable en $\hat{x}_0$ y γ es derivable en t_0, entonces $f \circ \gamma$ es derivable en t_0 y además*

$$(f \circ \gamma)'(t_0) = \nabla f(\hat{x}_0) \cdot \gamma'(t_0).$$

Demostración. Dado que la composición $f \circ \gamma$ es una función de $\mathbb{R}$ en $\mathbb{R}$, de acuerdo con la definición de la derivada de este tipo de funciones, debemos de mostrar que

$$\lim_{t \to t_0} \frac{(f \circ \gamma)(t) - (f \circ \gamma)(t_0)}{t - t_0}$$

existe.

De esta forma, lo primero que haremos será escribir de manera más "adecuada" a este cociente.

Nótese que, si $\gamma(t) - \gamma(t_0) \neq \hat{0}$, entonces

$$\begin{aligned}
&\frac{(f \circ \gamma)(t) - (f \circ \gamma)(t_0)}{t - t_0} \\
&= \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} \left(\frac{f(\gamma(t)) - f(\gamma(t_0)) - \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} \right. \\
&\quad \left. + \frac{\nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} \right) \\
&= \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} \left(\frac{f(\gamma(t)) - f(\gamma(t_0)) - \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} \right) \\
&\quad + \nabla f(\gamma(t_0)) \cdot \left(\frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right).
\end{aligned}$$

Ahora, obsérvese que se satisfacen los siguientes tres hechos:

1. como γ es derivable en t_0, γ es continua en t_0, es decir que

$$\lim_{t\to t_0} \gamma(t) = \gamma(t_0),$$

2. la expresión

$$\frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0}$$

está acotada en una vecindad (agujerada) de t_0 pues, como γ es derivable en t_0, se tiene que

$$\begin{aligned}\lim_{t\to t_0} \left| \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} \right| &= \lim_{t\to t_0} \left\| \frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right\| \\ &= \left\|\gamma'(t_0)\right\|,\end{aligned}$$

y

3. como f es derivable en $\hat{x}_0 = \gamma(t_0)$, entonces

$$\lim_{t\to t_0} \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} \left(\frac{f(\gamma(t)) - f(\gamma(t_0)) - \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} \right) = 0.$$

Finalmente, por el inciso 3 de la proposición 2.9, se tiene que

$$\begin{aligned}\lim_{t\to t_0} \nabla f(\gamma(t_0)) \cdot \left(\frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right) &= \lim_{t\to t_0} \nabla f(\gamma(t_0)) \cdot \lim_{t\to t_0} \frac{\gamma(t) - \gamma(t_0)}{t - t_0} \\ &= \nabla f(\gamma(t_0)) \cdot \gamma'(t_0),\end{aligned}$$

con lo cual obtendríamos el resultado deseado.

Sin embargo (y como el lector ya habrá notado), el argumento anterior sólo es correcto si $\gamma(t) - \gamma(t_0) \neq \hat{0}$. Para que el razonamiento anterior sea válido incluso en el caso en el que $\gamma(t) - \gamma(t_0) = \hat{0}$, nuevamente recurriremos a una función auxiliar que resuelve el problema.

Sea $\varphi : (a,b) \subset \mathbb{R} \to \mathbb{R}$ definida como

$$\varphi(t) = \begin{cases} \frac{f(\gamma(t)) - f(\gamma(t_0)) - \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} & \text{si } \gamma(t) - \gamma(t_0) \neq \hat{0} \\ 0 & \text{si } \gamma(t) - \gamma(t_0) = \hat{0} \end{cases}$$

Es fácil verificar que

$$\frac{(f \circ \gamma)(t) - (f \circ \gamma)(t_0)}{t - t_0} = \varphi(t) \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} + \nabla f(\gamma(t_0)) \cdot \left(\frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right)$$

para toda $t \in (a,b)$, $t \neq t_0$. Por otra parte, como el lector probará en el problema 16, φ es continua en t_0 de tal forma que

$$\lim_{t \to t_0} \varphi(t) \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} = 0,$$

de donde

$$\begin{aligned} &\lim_{t \to t_0} \frac{(f \circ \gamma)(t) - (f \circ \gamma)(t_0)}{t - t_0} \\ &\quad = \lim_{t \to t_0} \left(\varphi(t) \frac{\|\gamma(t) - \gamma(t_0)\|}{t - t_0} + \nabla f(\gamma(t_0)) \cdot \left(\frac{\gamma(t) - \gamma(t_0)}{t - t_0} \right) \right) \\ &\quad = \nabla f(\gamma(t_0)) \cdot \gamma'(t_0), \end{aligned}$$

que es lo que deseábamos demostrar. ■

Esta última proposición, además de proporcionarnos una excelente regla de derivación, tiene consecuencias de caracter práctico muy importantes cuya discusión ocupará un espacio grande.

Consideremos una función f de $\mathbb{R}^n$ en $\mathbb{R}$ y un conjunto de nivel $N_c(f) := \{\hat{x} \in U \subset \mathbb{R}^n \mid f(\hat{x}) = c\}$ de ésta. Un problema que resulta de interés es el de encontrar un vector que esté en la dirección "normal" al conjunto de nivel $N_c(f)$, en un punto específico de dicho conjunto. Lo que haremos a continuación será mostrar que si $\hat{x}_0$ pertenece al conjunto de nivel $N_c(f) := \{\hat{x} \in U \subset \mathbb{R}^n \mid f(\hat{x}) = c\}$, entonces $\nabla f(\hat{x}_0)$ es "normal" a $N_c(f)$ en $\hat{x}_0$ en el

sentido de que se satisface lo siguiente: para cualquier $\gamma : (a, b) \subset \mathbb{R} \to \mathbb{R}^n$ con las características de que $\gamma(t) \in N_c(f)$ para toda $t \in (a, b)$, γ es derivable en $t_0 \in (a, b)$ y $\gamma(t_0) = \hat{x}_0$, se cumple que

$$\nabla f(\gamma(t_0)) \cdot \gamma'(t_0) = 0.$$

Esta afirmación la dejaremos formulada como un corolario de la proposición 4.16 y su demostración se deja al lector.

Corolario 4.1 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ derivable en $\hat{x}_0 \in U$ tal que $\nabla f(\hat{x}_0) \neq \hat{0}$. Si $\hat{x}_0 \in N_c(f)$ (el conjunto de nivel c de f), entonces $\nabla f(\hat{x}_0)$ es "normal" a $N_c(f)$ en $\hat{x}_0$. Es decir que para cualquier $\gamma : (a, b) \subset \mathbb{R} \to \mathbb{R}^n$ tal que: $\gamma(t) \in N_c(f)$ para toda $t \in (a, b)$, γ es derivable en $t_0 \in (a, b)$ y $\gamma(t_0) = \hat{x}_0$, se tiene que*

$$\nabla f(\gamma(t_0)) \cdot \gamma'(t_0) = 0.$$

Como se ilustró en el capítulo 2, una buena cantidad de objetos geométricos muy conocidos (sobre todo en $\mathbb{R}^2$ y en $\mathbb{R}^3$) se pueden ver como conjuntos de nivel de una función de $\mathbb{R}^n$ en $\mathbb{R}$. Lo interesante del corolario anterior es que nos proporciona los elementos suficientes para definir (de forma muy sencilla) el concepto de recta tangente y plano tangente para el caso particular de conjuntos de nivel en $\mathbb{R}^2$ y en $\mathbb{R}^3$, respectivamente.

Definición 4.6 *Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ derivable en $\hat{x}_0 = (x_0, y_0) \in U$ tal que $\nabla f(\hat{x}_0) \neq \hat{0}$. Decimos que la recta (en $\mathbb{R}^2$) determinada por la ecuación cartesiana*

$$\frac{\partial f}{\partial x}(\hat{x}_0)\,(x - x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)\,(y - y_0) = 0$$

es la recta tangente en $\hat{x}_0$ del conjunto (o curva) de nivel $N_{f(\hat{x}_0)}(f)$.

Definición 4.7 *Sea $f : U \subset \mathbb{R}^3 \to \mathbb{R}$ derivable en $\hat{x}_0 = (x_0, y_0, z_0) \in U$ tal que $\nabla f(\hat{x}_0) \neq \hat{0}$. Decimos que el plano (en $\mathbb{R}^3$) determinado por la ecuación cartesiana*

$$\frac{\partial f}{\partial x}(\hat{x}_0)\,(x - x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)\,(y - y_0) + \frac{\partial f}{\partial z}(\hat{x}_0)\,(z - z_0) = 0$$

es el plano tangente en $\hat{x}_0$ del conjunto de nivel $N_{f(\hat{x}_0)}(f)$.

A continuación daremos un par de ejemplos en los que se muestra cómo se obtienen la recta y el plano tangente definidos anteriormente.

Ejemplo 4.10 *Consideremos:*

1. *la curva determinada por la ecuación*

$$(x^2+y^2-2x)^2=4(x^2+y^2),$$

la cual corresponde a la ya conocida cardioide.

Lo que ahora queremos hacer notar es que la cardioide es el conjunto de nivel 0 *de la función*

$$f(x,y)=(x^2+y^2-2x)^2-4(x^2+y^2),$$

que resultará muy conveniente si se quieren calcular rectas tangentes a este conjunto.

Para esta función se tiene que

$$\frac{\partial f}{\partial x}(x,y)=2(x^2+y^2-2x)\,(2x-2)-8x$$

y

$$\frac{\partial f}{\partial y}(x,y)=2(x^2+y^2-2x)\,(2y)-8y,$$

de modo que

$$\begin{aligned}&\nabla f(x,y)\\&=4\big((x^2+y^2-2x)\,(x-1)-2x,y(x^2+y^2-2x-2)\big)\\&=4\Big(\Big((x-1)^2+y^2-1\Big)\,(x-1)-2\,(x-1)-2,y\,\Big((x-1)^2+y^2-3\Big)\Big)\\&=4\Big((x-1)\,\Big((x-1)^2+y^2-3\Big)-2,y\,\Big((x-1)^2+y^2-3\Big)\Big).\end{aligned}$$

De esta forma, si $\nabla f(x_0,y_0)\neq(0,0)$*, este vector es normal a la curva (en el punto* (x_0,y_0)*) y la recta dada por la ecuación*

$$\nabla f(x_0,y_0)\cdot(x-x_0,y-y_0)=0,$$

o equivalentemente

$$\frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) = 0,$$

es la ecuación de la recta tangente a la curva (en el punto (x_0, y_0)*).*

Si en particular tomamos los puntos en que la cardioide intersecta al eje Y*, es decir los puntos* $(0, 2)$ *y* $(0, -2)$*, se tiene que*

$$\nabla f(0, 2) = (-16, 16) \qquad y \qquad \nabla f(0, -2) = (-16, -16).$$

Por lo tanto, la ecuación de las respectivas rectas tangentes están dadas por

$$-16x + 16(y - 2) = 0 \qquad o \qquad y = x + 2$$

y

$$-16x - 16(y + 2) = 0 \qquad o \qquad y = -x - 2,$$

que son las que ya habíamos obtenido en el capítulo 3.

2. *Sea* $S \subset \mathbb{R}^3$ *el conjunto definido como*

$$S = \left\{ (x, y, z) \in \mathbb{R}^3 \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \right\},$$

el cual corresponde a un elipsoide con centro en el $(0, 0, 0)$ *(ver figura 4.8).*

Como en el caso de la cardioide, este conjunto tampoco se puede obtener (completo) como la gráfica de una función de $\mathbb{R}^2$ *en* $\mathbb{R}$*.*

Aun cuando algunas partes de este conjunto sí se pueden ver como una de estas gráficas, en términos de los cálculos que hay que realizar (usando la ecuación 4.16), resulta mucho más fácil observar que se puede obtener como el conjunto de nivel 1 *de la función*

$$f(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}.$$

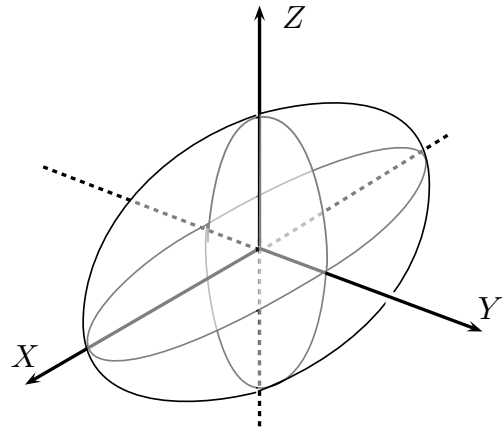


Figura 4.8: El elipsoide determinado por la ecuación $\frac{x^2}{8}+\frac{y^2}{3}+\frac{z^2}{5}=1$, análogo al del ejemplo 4.10.

Como

$$\nabla f(x,y,z) = 2\left(\frac{x}{a^2},\frac{y}{b^2},\frac{z}{c^2}\right),$$

si (x_0,y_0,z_0) *es cualquier punto del elipsoide, entonces* $\nabla f(x_0,y_0,z_0)\neq(0,0,0)$ *y por lo tanto la ecuación del plano tangente en este punto estará dada por*

$$\begin{aligned}\nabla f(x_0,y_0,z_0)\cdot(x-x_0,y-y_0,z-z_0) &= 0\\ &= \left(\frac{x_0}{a^2},\frac{y_0}{b^2},\frac{z_0}{c^2}\right)\cdot(x-x_0,y-y_0,z-z_0).\end{aligned}$$

Al simplificarla, se reduce a la ecuación

$$\left(\frac{x_0}{a^2}\right)x+\left(\frac{y_0}{b^2}\right)y+\left(\frac{z_0}{c^2}\right)z=1.$$

Como el lector habrá notado, en el ejemplo anterior se hace énfasis en que los conjuntos involucrados no se pueden obtener como la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$, para el primer inciso, o de una función de $\mathbb{R}^2$ en $\mathbb{R}$,

para el segundo. La razón de ello es que el concepto de recta tangente y plano tangente ya se tenían definidos para este tipo de objetos, pero dada la incapacidad de verlos de esta forma, se quiso resaltar la "conveniencia" de poderlos obtener como un conjunto de nivel de cierta función, para así poder aplicar las definiciones 4.6 y 4.7.

Pero si bien es cierto que no todo conjunto de nivel en $\mathbb{R}^2$ se puede ver como la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$, ni todo conjunto de nivel en $\mathbb{R}^3$ se puede ver como la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$, lo recíproco sí se cumple (problema 10 del capítulo 2). Pues bien, en los problemas 10 y 11 de este capítulo pedimos al lector que pruebe que los conceptos de recta y plano tangente dados en las definiciones 4.6 y 4.7, coinciden con los conceptos correspondientes para los de gráfica de una función.

Más adelante, en el capítulo 5, probaremos un importante resultado (el Teorema de la Función Implícita) del cual podremos deducir, bajo ciertas hipótesis, que un conjunto de nivel en $\mathbb{R}^2$ (o en $\mathbb{R}^3$) sí se puede obtener, cuando menos "localmente", como la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$ (de $\mathbb{R}^2$ en $\mathbb{R}$). Apoyados en este teorema, se podrá probar entonces que la recta y plano tangente definidos en 4.6 y 4.7 también se pueden obtener como la recta (o plano) tangente a la gráfica de cierta función de $\mathbb{R}$ en $\mathbb{R}$ (o de $\mathbb{R}^2$ en $\mathbb{R}$), respectivamente.

Otra interpretación del resultado de la proposición 4.16 que tendrá un uso muy importante es el siguiente: si $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $\gamma : (a, b) \subset \mathbb{R} \to \mathbb{R}^n$, con $\gamma(a, b) \subset U$, y $t_0 \in (a, b)$ son tales que $\gamma(t_0) = \hat{x}_0$, es un hecho que la composición $(f \circ \gamma)(t) = f(\gamma(t))$ nos da los valores de f sobre los puntos de la curva recorrida por la función γ.

Con base en lo anterior, es fácil convencerse de que la derivada de esta composición en t_0 se puede interpretar como la razón de cambio de f en $\hat{x}_0$ cuando nos aproximamos a este punto por la curva descrita por γ. Aquí es importante destacar que resulta "natural" que esta razón de cambio se vea afectada por la velocidad con la que γ se aproxima al punto $\hat{x}_0$, lo que queda evidenciado en la fórmula

$$(f \circ \gamma)'(t_0) = \nabla f(\gamma(t_0)) \cdot \gamma'(t_0) \tag{4.27}$$

De esta manera, si lo que deseamos es interpretar a $(f \circ \gamma)'(t_0)$ como la razón de cambio de f en $\hat{x}_0$ (cuando nos acercamos a este punto por la curva descrita por la función γ), es importante que γ sea tal que $\gamma'(t_0) \neq \hat{0}$. De hecho, si suponemos que γ es una parametrización por longitud de arco, es decir que $\|\gamma'(t)\| = 1$ para cada $t \in (a, b)$, para que no haya problema con la velocidad (o para ser más exactos, con la rapidez) con la que nos aproximamos a $\hat{x}_0$, la derivada de $f \circ \gamma$ en t_0 no es más que la derivada direccional de f en $\hat{x}_0$, en la dirección del vector $\gamma'(t_0)$ (que ahora sería de norma 1).

En efecto, si ahora usamos la notación $Df(\hat{x}_0)$ para referirnos a la derivada de f en $\hat{x}_0$, usando la proposición 4.10, se tiene que

$$\begin{aligned}(f \circ \gamma)'(t_0) &= \nabla f(\gamma(t_0)) \cdot \gamma'(t_0) \\ &= Df(\hat{x}_0)(\gamma'(t_0)) \\ &= D_{\gamma'(t_0)} f(\hat{x}_0).\end{aligned}$$

Más aún, si sólo se tiene que $\gamma'(t_0) \neq \hat{0}$ (aunque este vector no sea de norma 1), la identidad 4.27 se puede escribir como

$$\begin{aligned}(f \circ \gamma)'(t_0) &= \nabla f(\gamma(t_0)) \cdot \gamma'(t_0) \\ &= \left\|\gamma'(t_0)\right\| \left(\nabla f(\gamma(t_0)) \cdot \left(\frac{\gamma'(t_0)}{\|\gamma'(t_0)\|}\right)\right) \\ &= \left\|\gamma'(t_0)\right\| Df(\hat{x}_0) \left(\frac{\gamma'(t_0)}{\|\gamma'(t_0)\|}\right) \\ &= \left\|\gamma'(t_0)\right\| D_{\frac{\gamma'(t_0)}{\|\gamma'(t_0)\|}} f(\hat{x}_0). \qquad (4.28)\end{aligned}$$

Que la razón de cambio de una función f sobre la curva descrita por una función γ en un punto $\hat{x}_0$ sea igual a la derivada direccional de f en $\hat{x}_0$, en la dirección de $\gamma'(t_0)$ (normalizado) multiplicada por la rapidez con que γ se aproxima a $\hat{x}_0$, no hace más que confirmar que el concepto de derivada es un concepto "local" (simplemente recuerde el lector que, cerca del punto $\hat{x}_0 = \gamma(t_0)$, γ y su recta tangente se "parecen" mucho).

Ilustremos la discusión anterior con el siguiente

Ejemplo 4.11 *Sean* $f : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ *definida como*

$$f(x,y) = \frac{xy}{x^2+y^2}$$

y $\gamma : \mathbb{R} \to \mathbb{R}^2$ *definida como* $\gamma(t) = (r\cos(t), r\,\mathrm{sen}(t))$.

Nótese que

$$\begin{aligned}(f\circ\gamma)(t) &= f(\gamma(t))\\ &= \frac{(r\cos(t))\,(r\,\mathrm{sen}(t))}{(r\cos(t))^2+(r\,\mathrm{sen}(t))^2}\\ &= \frac{\cos(t)\,\mathrm{sen}(t)}{\cos^2(t)+\mathrm{sen}^2(t)}\\ &= \cos(t)\,\mathrm{sen}(t),\end{aligned}$$

de modo que

$$\begin{aligned}(f\circ\gamma)'(t) &= \cos(t)\,\mathrm{sen}'(t)+\cos'(t)\,\mathrm{sen}(t)\\ &= \cos^2(t)-\mathrm{sen}^2(t).\end{aligned}$$

Por otra parte, se tiene que

$$\begin{aligned}\nabla f(x,y) &= \left(\frac{\partial f}{\partial x}(x,y), \frac{\partial f}{\partial y}(x,y)\right)\\ &= \left(\frac{y\left(x^2+y^2\right)-2x\,(xy)}{\left(x^2+y^2\right)^2}, \frac{x\left(x^2+y^2\right)-2y\,(xy)}{\left(x^2+y^2\right)^2}\right)\\ &= \left(\frac{y\left(y^2-x^2\right)}{\left(x^2+y^2\right)^2}, \frac{x\left(x^2-y^2\right)}{\left(x^2+y^2\right)^2}\right)\end{aligned}$$

y

$$\gamma'(t) = (-r\,\mathrm{sen}(t), r\cos(t)).$$

Por tanto, de acuerdo con la identidad 4.27 se debe tener que

$$(f\circ\gamma)'(t) = \nabla f(\gamma(t))\cdot\gamma'(t)$$

de tal forma que, como

$$\begin{aligned}
&\nabla f(\gamma(t)) \\
&= \left(\frac{r\,\mathrm{sen}(t)\left((r\,\mathrm{sen}(t))^2 - (r\cos(t))^2\right)}{\left((r\cos(t))^2 + (r\,\mathrm{sen}(t))^2\right)^2}, \frac{r\cos(t)\left((r\cos(t))^2 - (r\,\mathrm{sen}(t))^2\right)}{\left((r\cos(t))^2 + (r\,\mathrm{sen}(t))^2\right)^2} \right) \\
&= \frac{1}{r}\left(\mathrm{sen}(t)\left(\mathrm{sen}^2(t) - \cos^2(t)\right), \cos(t)\left(\cos^2(t) - \mathrm{sen}^2(t)\right)\right) \\
&= \left(\cos^2(t) - \mathrm{sen}^2(t)\right)\left(-\mathrm{sen}(t), \cos(t)\right),
\end{aligned}$$

entonces

$$\begin{aligned}
(f \circ \gamma)'(t) &= \nabla f(\gamma(t)) \cdot \gamma'(t) \\
&= \left(\cos^2(t) - \mathrm{sen}^2(t)\right)\left(-\mathrm{sen}(t), \cos(t)\right) \cdot \left(-\mathrm{sen}(t), \cos(t)\right) \\
&= \cos^2(t) - \mathrm{sen}^2(t),
\end{aligned}$$

lo cual coincide con el primer cálculo que hicimos.

La identidad

$$(f \circ \gamma)'(t_0) = \left\|\gamma'(t_0)\right\| D_{\frac{\gamma'(t_0)}{\|\gamma'(t_0)\|}} f(\hat{x}_0)$$

(bajo el supuesto de que $\gamma'(t_0) \neq \hat{0}$) tendrá implicaciones prácticas muy importantes, sobre todo en lo relacionado con el cálculo de la derivada de una función f cuando ésta esté descrita en términos de coordenadas diferentes a las cartesianas, como las que mencionamos en el capítulo 1. Justo esto es lo que nos disponemos a desarrollar en la siguiente sección.

4.3.3. La derivada en otras coordenadas

En el capítulo 1 introdujimos, para el caso particular de los puntos o vectores del plano o del espacio, otros sistemas de coordenadas diferentes de las coordenadas cartesianas. En ese capítulo ya describimos las características

más importantes de esas otras coordenadas, de tal forma que aquí las vamos a usar con toda libertad.

Nuestro objetivo en esta sección es encontrar una representación de la derivada de una función que está dada en términos de algunas de estas coordenadas. Para ello, es importante tener presentes los siguientes dos hechos:

1. dado que la derivada es una función lineal, para tener una representación de ella es suficiente conocer sus valores en los elementos de una base ortonormal de $\mathbb{R}^2$ o $\mathbb{R}^3$ (dependiendo de dónde esté contenido el dominio de la función con la que estemos tratando), y

2. si nuestra función está expresada en términos de algunas de estas coordenadas (polares, cilíndricas o esféricas), los únicos cálculos que estaremos en condiciones de hacer, consistirán en "derivar" estas expresiones con respecto a las variables en las que estén escritas.

Empezaremos por analizar el segundo punto, y para ello supondremos que tenemos una función $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ que está expresada en términos de las coordenadas polares (ρ, θ) de cada punto $\hat{x} \in U$.

Derivar la expresión que define a la función f con respecto a una de sus variables, por ejemplo ρ, significará que en dicha expresión sólo consideraremos como variable a ρ, y a la variable θ la trataremos como una constante. En términos más elementales, es decir, recordando que todo proceso de derivación con respecto de una variable consiste en el cálculo de un límite (independientemente de que para obtener estas derivadas hagamos uso de algunas de las reglas de derivación que ya conocemos), lo que estaremos haciendo es obtener el siguiente límite:

$$\lim_{h \to 0} \frac{f(\rho + h, \theta) - f(\rho, \theta)}{h}. \tag{4.29}$$

Sin duda el cociente que aparece en el límite anterior representa una razón de cambio de la función f, y lo que analizaremos con más cuidado es sobre qué puntos del dominio de f se está calculando esta razón de cambio.

Para tener una idea más precisa de qué tipo de puntos son aquellos cuyas coordenadas polares son de la forma $(\rho + h, \theta)$, recurriremos a las fórmulas

de conversión entre coordenadas polares y coordenadas cartesianas que dedujimos en el capítulo 1. De esta forma, si llamamos $\hat{x} \in U$ al punto cuyas coordenadas polares son (ρ, θ), tenemos que las coordenadas cartesianas de $\hat{x}$ están dadas por la pareja $(\rho\cos(\theta), \rho\,\mathrm{sen}(\theta))$, y si llamamos $\hat{x}_h \in U$ al punto cuyas coordenadas polares son $(\rho + h, \theta)$, sus coordenadas cartesianas estarán dadas por $((\rho + h)\cos(\theta), (\rho + h)\,\mathrm{sen}(\theta))$.

Nótese entonces que

$$\begin{aligned}\hat{x}_h &= ((\rho + h)\cos(\theta), (\rho + h)\,\mathrm{sen}(\theta)) \\ &= (\rho\cos(\theta), \rho\,\mathrm{sen}(\theta)) + h(\cos(\theta), \mathrm{sen}(\theta)) \\ &= \hat{x} + h(\cos(\theta), \mathrm{sen}(\theta)),\end{aligned}$$

lo que significa que $\hat{x}_h$ siempre es un punto que está en la recta que pasa por $\hat{x}$, en la dirección del vector $(\cos(\theta), \mathrm{sen}(\theta))$ (justo la que hace un ángulo θ con el eje polar) (ver figura 4.9).

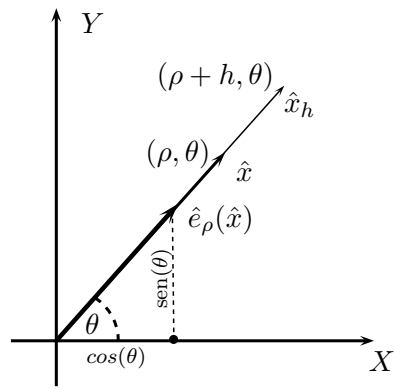


Figura 4.9: Si un vector $\hat{x}$ tiene coordenadas polares (ρ, θ), entonces el vector $\hat{x}_h$ de coordenadas polares $(\rho + h, \theta)$ está en la misma dirección que $\hat{x}$ (si ρ y $\rho + h$ tienen el mismo signo), y que el vector unitario $\hat{e}_\rho(\hat{x})$ de coordenadas cartesianas $(cos(\theta), \mathrm{sen}(\theta))$.

En virtud de lo anterior, si denotamos por $\hat{e}_\rho(\hat{x})$ al vector de coordenadas (cartesianas) $(\cos(\theta), \mathrm{sen}(\theta))$, tenemos que el límite 4.29 se reescribe como

$$\lim_{h\to 0}\frac{f(\rho+h,\theta)-f(\rho,\theta)}{h}=\lim_{h\to 0}\frac{f(\hat{x}+h\hat{e}_\rho(\hat{x}))-f(\hat{x})}{h},$$

de tal forma que, por la definición 4.1 y la proposición 4.10, se tiene que

$$\begin{aligned}\lim_{h\to 0}\frac{f(\rho+h,\theta)-f(\rho,\theta)}{h}&=\lim_{h\to 0}\frac{f(\hat{x}+h\hat{e}_\rho(\hat{x}))-f(\hat{x})}{h}\\&=D_{\hat{e}_\rho(\hat{x})}f(\hat{x})\\&=Df(\hat{x})(\hat{e}_\rho(\hat{x})),\end{aligned}$$

o lo que es lo mismo, que

$$Df(\hat{x})(\hat{e}_\rho(\hat{x}))=\lim_{h\to 0}\frac{f(\rho+h,\theta)-f(\rho,\theta)}{h}. \tag{4.30}$$

Esta última identidad es muy afortunada, ya que nos dice que el proceso de derivación con respecto de la variable ρ que describimos al inicio de esta discusión, coincide con ser la evaluación de la derivada de f en $\hat{x}$ ($Df(\hat{x})$) en el vector unitario $\hat{e}_\rho(\hat{x})$. Esto es sin duda de mucha ayuda para lograr nuestro objetivo (el de obtener una expresión de $Df(\hat{x})$, como lo mencionamos al inicio de esta sección).

Seguramente el lector estará de acuerdo en que ahora lo que hay que hacer es la otra derivada, es decir, derivar la expresión que define a f con respecto de la variable θ (tratando a la variable ρ como constante), y esperar que esta otra derivada también coincida con ser la evaluación de la derivada de f en $\hat{x}$ ($Df(\hat{x})$) en algún otro vector unitario, que junto con el vector $\hat{e}_\rho(\hat{x})$, forme una base ortonormal de $\mathbb{R}^2$ (¡soñar no cuesta nada!). Aun cuando este sueño no se cumplirá completo, ¡lograremos despertar con una solución al problema que nos planteamos resolver en esta sección! ¡Manos a la obra!

Derivar la expresión que define a f con respecto de la variable θ (tratando a la variable ρ como constante), nuevamente se traduce en calcular el siguiente límite:

$$\lim_{h\to 0}\frac{f(\rho,\theta+h)-f(\rho,\theta)}{h}. \tag{4.31}$$

Como en el caso anterior, el cociente que aparece en este límite también representa una razón de cambio de la función f, y lo que haremos otra vez, será analizar sobre qué puntos del dominio de f se está calculando esta razón de cambio, para lo cual recurriremos nuevamente a las fórmulas de conversión entre coordenadas polares y coordenadas cartesianas.

Si, como hicimos antes, llamamos $\hat{x}$ al punto cuyas coordenadas polares son (ρ, θ), y $\hat{x}_h$ ahora al de coordenadas polares $(\rho, \theta + h)$, tendremos que las correspondientes coordenadas cartesianas serán $(\rho\cos(\theta), \rho\,\mathrm{sen}(\theta))$ y $(\rho\cos(\theta+h), \rho\,\mathrm{sen}(\theta+h))$, respectivamente. Con base en estas coordenadas cartesianas, es fácil notar ahora que $\hat{x}$ y $\hat{x}_h$ son puntos que pertenecen a la misma circunferencia (la de radio ρ con centro en el origen), y que si h es un número "cercano" a 0, entonces $\hat{x}_h$ es un punto "cercano" a $\hat{x}$ (ver figura 4.10).

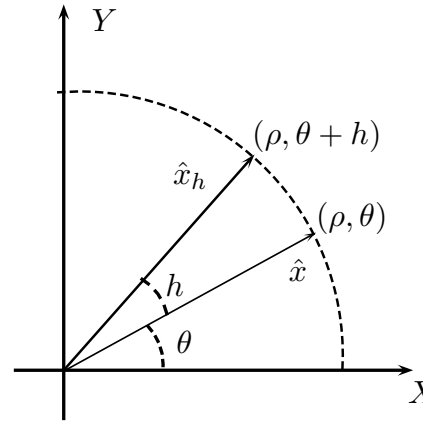


Figura 4.10: Si un vector $\hat{x}$ tiene coordenadas polares (ρ, θ), entonces el vector $\hat{x}_h$ de coordenadas polares $(\rho, \theta + h)$ está en la misma circunferencia (la de radio ρ) que $\hat{x}$.

De esta forma, la derivada con respecto de la variable θ de la función f (que se obtiene al calcular el límite 4.31) se puede interpretar como el cálculo de la razón de cambio de la función f cuando nos aproximamos a $\hat{x}$ por puntos de la misma circunferencia que pasa por este punto.

Por lo tanto, si parametrizamos esta circunferencia por la función $\gamma : \mathbb{R} \to \mathbb{R}^2$ definida como

$$\gamma(h) = (\rho\cos(\theta + h), \rho\,\mathrm{sen}(\theta + h)),$$

el límite 4.31 también se puede reescribir como

$$\begin{aligned}\lim_{h\to 0} \frac{f(\rho, \theta + h) - f(\rho, \theta)}{h} &= \lim_{h\to 0} \frac{f(\gamma(h)) - f(\gamma(0))}{h} \\ &= (f \circ \gamma)'(0).\end{aligned}$$

Como seguramente el lector recordará, la proposición 4.16 nos asegura que

$$\begin{aligned}(f \circ \gamma)'(0) &= \nabla f(\gamma(0)) \cdot \gamma'(0) \\ &= \nabla f(\hat{x}) \cdot (-\rho\,\mathrm{sen}(\theta), \rho\cos(\theta)) \\ &= \rho\nabla f(\hat{x}) \cdot (-\mathrm{sen}(\theta), \cos(\theta)) \\ &= \rho Df(\hat{x})(-\mathrm{sen}(\theta), \cos(\theta)),\end{aligned}$$

de tal forma que si ahora denotamos por $\hat{e}_\theta(\hat{x})$ al vector de coordenadas cartesianas $(-\mathrm{sen}(\theta), \cos(\theta))$ (¡que es de norma 1 y perpendicular al vector $\hat{e}_\rho(\hat{x})$!), tenemos que

$$\lim_{h\to 0} \frac{f(\rho, \theta + h) - f(\rho, \theta)}{h} = \rho Df(\hat{x})(\hat{e}_\theta(\hat{x})).$$

Aun cuando esta última identidad no nos asegura que la derivada con respecto de la variable θ de la función f sea exactamente igual a la derivada de f en $\hat{x}$ $(Df(\hat{x}))$ evaluada en algún vector unitario (como sucedió en el caso anterior), si $\rho \neq 0$, concluimos que

$$Df(\hat{x})(\hat{e}_\theta(\hat{x})) = \frac{1}{\rho} \lim_{h\to 0} \frac{f(\rho, \theta + h) - f(\rho, \theta)}{h}. \tag{4.32}$$

Las identidades 4.30 y 4.32 nos proporcionan la solución al problema que nos planteamos al inicio de esta sección (para el caso de las coordenadas

polares): cómo encontrar una representación de la derivada de una función que está dada en términos de este tipo de coordenadas.

Es fundamental que el lector tenga claro la forma en que quedó resuelto este problema. Como mencionamos en el punto número 1 al inicio de esta sección, dado que la derivada de una función es una función lineal, ésta queda completamente determinada si se conocen sus valores en los elementos de una base ortonormal (en este caso de $\mathbb{R}^2$). Y las identidades 4.30 y 4.32 justo nos proporcionan estos valores para la base ortonormal formada por los vectores $\hat{e}_\rho(\hat{x})$ y $\hat{e}_\theta(\hat{x})$, que sin duda forman una de estas bases.

Lo que es importante destacar es que estos vectores dependen del punto en el que estemos "calculando" la derivada de f (lo que por cierto, explica la notación que usamos para representarlos); si cambiamos de punto, cambiamos de base en la cual estamos representando a la función lineal $Df(\hat{x})$ (ver figura 4.11).

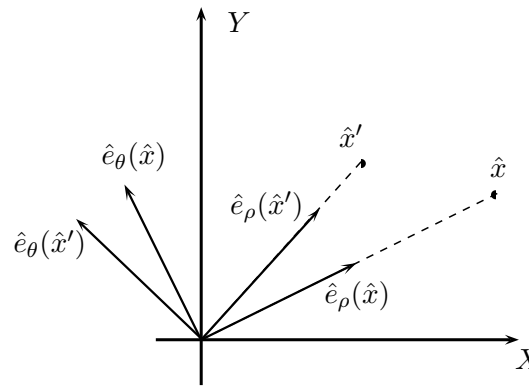


Figura 4.11: Cuando una función f está dada en términos de coordenadas polares, la derivada de f se expresa en una base ortonormal que depende del punto en donde se evalue, $\{\hat{e}_\rho(\hat{x}), \hat{e}_\theta(\hat{x})\}$ para el punto $\hat{x}$, y $\{\hat{e}_\rho(\hat{x}'), \hat{e}_\theta(\hat{x}')\}$ para el punto $\hat{x}'$.

El otro aspecto importante de la solución que encontramos a nuestro problema está en la forma en que están dados los valores de la derivada de f en $\hat{x}$ ($Df(\hat{x})$) en los vectores $\hat{e}_\rho(\hat{x})$ y $\hat{e}_\theta(\hat{x})$: calculando lo único que

podíamos calcular (como lo mencionamos en el punto 2 al inicio de esta sección), la derivada con respecto a las variables ρ y θ de la expresión que define a f, y que se obtienen por medio de los límites que aparecen en las identidades 4.30 y 4.32.

Estas "derivadas parciales", aunque análogas a las que definimos en 4.2 no son del mismo tipo en virtud de que no siempre resultan ser una derivada direccional. Por esta razón, lo siguiente que haremos será definir este otro tipo de derivadas parciales.

Definición 4.8 *Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ una función que está expresada en términos de las coordenadas polares (ρ, θ) de cada punto $\hat{x} \in U$. Si $\hat{x}_0 \in U$ tiene coordenadas polares (ρ_0, θ_0), definimos:*

1. *la derivada parcial de f con respecto de ρ en $\hat{x}_0$, que denotamos por $\frac{\partial f}{\partial \rho}(\hat{x}_0)$, como*

$$\frac{\partial f}{\partial \rho}(\hat{x}_0) := \lim_{h \to 0} \frac{f(\rho_0 + h, \theta_0) - f(\rho_0, \theta_0)}{h}$$

2. *la derivada parcial de f con respecto de θ en $\hat{x}_0$, que denotamos por $\frac{\partial f}{\partial \theta}(\hat{x}_0)$, como*

$$\frac{\partial f}{\partial \theta}(\hat{x}_0) := \lim_{h \to 0} \frac{f(\rho_0, \theta_0 + h) - f(\rho_0, \theta_0)}{h}.$$

Con base en estas definiciones podemos resumir lo obtenido hasta ahora de la siguiente manera: si $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ es una función que está expresada en términos de las coordenadas polares (ρ, θ) de cada punto $\hat{x} \in U$, y $\hat{x}_0 \in U$, $\hat{x}_0 \neq (0, 0)$, tiene coordenadas polares (ρ_0, θ_0), la derivada de f en $\hat{x}_0$ $(Df(\hat{x}_0))$ está representada en la base ortonormal formada por los vectores

$$\hat{e}_\rho(\hat{x}_0) := (\cos(\theta_0), \operatorname{sen}(\theta_0))$$
$$\hat{e}_\theta(\hat{x}_0) := (-\operatorname{sen}(\theta_0), \cos(\theta_0)),$$

por la matriz (de 1×2)

$$\left[\begin{array}{cc} \frac{\partial f}{\partial \rho}(\hat{x}_0) & \frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0) \end{array} \right]$$

o por el vector

$$\left(\frac{\partial f}{\partial \rho}(\hat{x}_0), \frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0) \right).$$

Esto significa que

$$\nabla f(\hat{x}_0) = \frac{\partial f}{\partial \rho}(\hat{x}_0)\hat{e}_\rho(\hat{x}_0) + \frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0)\hat{e}_\theta(\hat{x}_0),$$

es decir, que $\frac{\partial f}{\partial \rho}(\hat{x}_0)$ y $\frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0)$ son las coordenadas del gradiente de f en $\hat{x}_0$ ($\nabla f(\hat{x}_0)$) en la base $\{\hat{e}_\rho(\hat{x}_0), \hat{e}_\theta(\hat{x}_0)\}$.

Por otra parte, de acuerdo con las identidades 4.19 y 4.18, como

$$\begin{aligned} \hat{e}_\rho(\hat{x}_0) &= (\cos(\theta_0), \operatorname{sen}(\theta_0)) \\ &= \cos(\theta_0)\hat{e}_1 + \operatorname{sen}(\theta_0)\hat{e}_2 \end{aligned}$$

y

$$\begin{aligned} \hat{e}_\theta(\hat{x}_0) &= (-\operatorname{sen}(\theta_0), \cos(\theta_0)) \\ &= -\operatorname{sen}(\theta_0)\hat{e}_1 + \cos(\theta_0)\hat{e}_2, \end{aligned}$$

se tiene que

$$\begin{aligned} &\left[\begin{array}{cc} \frac{\partial f}{\partial x}(\hat{x}_0) & \frac{\partial f}{\partial y}(\hat{x}_0) \end{array} \right] \\ &\qquad = \left[\begin{array}{cc} \frac{\partial f}{\partial \rho}(\hat{x}_0) & \frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0) \end{array} \right] \left[\begin{array}{cc} \cos(\theta_0) & \operatorname{sen}(\theta_0) \\ -\operatorname{sen}(\theta_0) & \cos(\theta_0) \end{array} \right] \end{aligned} \tag{4.33}$$

y

$$\begin{aligned} &\left[\begin{array}{cc} \frac{\partial f}{\partial \rho}(\hat{x}_0) & \frac{1}{\rho_0}\frac{\partial f}{\partial \theta}(\hat{x}_0) \end{array} \right] \\ &\qquad = \left[\begin{array}{cc} \frac{\partial f}{\partial x}(\hat{x}_0) & \frac{\partial f}{\partial y}(\hat{x}_0) \end{array} \right] \left[\begin{array}{cc} \cos(\theta_0) & -\operatorname{sen}(\theta_0) \\ \operatorname{sen}(\theta_0) & \cos(\theta_0) \end{array} \right]. \end{aligned} \tag{4.34}$$

De estas últimas identidades de matrices se concluye que

$$\frac{\partial f}{\partial x}(\hat{x}_0) = \cos(\theta_0)\frac{\partial f}{\partial \rho}(\hat{x}_0) - \frac{1}{\rho_0}\operatorname{sen}(\theta_0)\frac{\partial f}{\partial \theta}(\hat{x}_0)$$
$$\frac{\partial f}{\partial y}(\hat{x}_0) = \operatorname{sen}(\theta_0)\frac{\partial f}{\partial \rho}(\hat{x}_0) + \frac{1}{\rho_0}\cos(\theta_0)\frac{\partial f}{\partial \theta}(\hat{x}_0)$$

y

$$\frac{\partial f}{\partial \rho}(\hat{x}_0) = \cos(\theta_0)\frac{\partial f}{\partial x}(\hat{x}_0) + \operatorname{sen}(\theta_0)\frac{\partial f}{\partial y}(\hat{x}_0)$$
$$\frac{\partial f}{\partial \theta}(\hat{x}_0) = -\rho_0\operatorname{sen}(\theta_0)\frac{\partial f}{\partial x}(\hat{x}_0) + \rho_0\cos(\theta_0)\frac{\partial f}{\partial y}(\hat{x}_0),$$

en donde recuerde que (ρ_0, θ_0) son las coordenadas polares de $\hat{x}_0$.

Con relación a la condición que impusimos de que $\hat{x}_0 \neq (0,0)$, como el lector ya sabrá, las coordenadas polares del origen son cualquier pareja de la forma $(0, \theta)$, en donde θ puede ser cualquier valor. Este hecho imposibilita la elección de la base ortonormal $\{\hat{e}_\rho(\hat{x}_0), \hat{e}_\theta(\hat{x}_0)\}$, lo que a su vez (en algunos casos) se ve reflejado en la imposibilidad de calcular el lado derecho de la identidad 4.32. Esto es lo que nos hace imponer la condición de que $\hat{x}_0 \neq (0,0)$.

El lector seguramente estará de acuerdo en que lo siguiente que habrá que hacer es un ejemplo que ilustre todo lo desarrollado en esta sección. Y eso es justo lo que haremos.

Ejemplo 4.12 *Considere la función $f : \mathbb{R}^2 \to \mathbb{R}$, que está expresada en términos de las coordenadas polares (ρ, θ) de cada punto $\hat{x} \in \mathbb{R}^2$ por*

$$f(\rho, \theta) = (\rho^2 - 2\rho\cos(\theta))^2 - 4\rho^2.$$

En este caso, tenemos que

$$\frac{\partial f}{\partial \rho}(\rho, \theta) = 2(\rho^2 - 2\rho\cos(\theta))\,(2\rho - 2\cos(\theta)) - 8\rho$$

y

$$\frac{\partial f}{\partial \theta}(\rho, \theta) = 2(\rho^2 - 2\rho\cos(\theta))\,(2\rho\operatorname{sen}(\theta))\,,$$

de tal forma que la derivada de f en un punto $\hat{x} \neq \hat{0}$ cuyas coordenadas polares sean (ρ, θ), es la función lineal $Df(\hat{x})$ cuya matriz asociada (de 1×2) en la base ortonormal $\{\hat{e}_\rho(\hat{x}), \hat{e}_\theta(\hat{x})\}$ es

$$\begin{bmatrix} \frac{\partial f}{\partial \rho}(\rho,\theta) & \frac{1}{\rho}\frac{\partial f}{\partial \theta}(\rho,\theta) \end{bmatrix}$$
$$= \begin{bmatrix} 2(\rho^2 - 2\rho\cos(\theta))(2\rho - 2\cos(\theta)) - 8\rho & 4\operatorname{sen}(\theta)(\rho^2 - 2\rho\cos(\theta)) \end{bmatrix},$$

o la determinada por el vector que tiene coordenadas (también en la base $\{\hat{e}_\rho(\hat{x}), \hat{e}_\theta(\hat{x})\}$)

$$\left(2(\rho^2 - 2\rho\cos(\theta))(2\rho - 2\cos(\theta)) - 8\rho, 4\operatorname{sen}(\theta)(\rho^2 - 2\rho\cos(\theta))\right).$$

Es decir, que

$$\nabla f(\rho,\theta) = \left(2(\rho^2 - 2\rho\cos(\theta))(2\rho - 2\cos(\theta)) - 8\rho, 4\operatorname{sen}(\theta)(\rho^2 - 2\rho\cos(\theta))\right).$$

Si en particular tomamos el punto $\hat{x}_0$, cuyas coordenadas polares son $(2, \pi/2)$, tenemos que las coordenadas del gradiente de f en $\hat{x}_0$ en la base $\{\hat{e}_\rho(\hat{x}_0), \hat{e}_\theta(\hat{x}_0)\}$ están dadas por

$$\nabla f(\hat{x}_0) = (16, 16).$$

Ahora, observe el lector que la función f de este ejemplo es la misma función que dimos en el inciso 1 del ejemplo 4.10 (basta con que recurra a las fórmulas de cambio de coordenadas para que compruebe este hecho). Nótese también, que el punto $\hat{x}_0$ que tiene coordenadas polares $(2, \pi/2)$, tiene coordenadas cartesianas $(0, 2)$ (en la base canónica $\{\hat{e}_1, \hat{e}_2\}$).

De esta forma, como

$$\begin{aligned} \hat{e}_\rho(\hat{x}_0) &= (\cos(\pi/2), \operatorname{sen}(\pi/2)) \\ &= (0, 1) \\ &= 0\hat{e}_1 + 1\hat{e}_2 \end{aligned}$$

y

$$\begin{aligned} \hat{e}_\theta(\hat{x}_0) &= (-\operatorname{sen}(\pi/2), \cos(\pi/2)) \\ &= (-1, 0) \\ &= -1\hat{e}_1 + 0\hat{e}_2, \end{aligned}$$

de acuerdo con la identidad 4.19, se debe tener que

$$\begin{aligned}\left[\begin{array}{cc} \frac{\partial f}{\partial x}(\hat{x}_0) & \frac{\partial f}{\partial y}(\hat{x}_0) \end{array}\right] &= \left[\begin{array}{cc} 16 & 16 \end{array}\right] \left[\begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array}\right] \\ &= \left[\begin{array}{cc} -16 & 16 \end{array}\right].\end{aligned}$$

Lo anterior es equivalente a que las coordenadas de $\nabla f(\hat{x}_0)$, en la base canónica $\{\hat{e}_1, \hat{e}_2\}$, deben estar dadas por

$$\begin{aligned}\nabla f(\hat{x}_0) &= \left(\frac{\partial f}{\partial x}(\hat{x}_0), \frac{\partial f}{\partial y}(\hat{x}_0)\right) \\ &= (-16, 16)\,,\end{aligned}$$

lo cual coincide con lo obtenido en el ejemplo mencionado.

Aun cuando hicimos el cálculo de la derivada sólo para funciones de $\mathbb{R}^2$ en $\mathbb{R}$ expresadas en términos de coordenadas polares, para funciones de $\mathbb{R}^3$ en $\mathbb{R}$ expresadas en términos de coordenadas cilíndricas o esféricas, el análisis es completamente análogo al anterior y lo dejaremos como un par de problemas para el lector.

4.4. Derivadas direccionales de orden superior

Si para una función $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y un vector unitario $\hat{u} \in \mathbb{R}^n$ fijo, se tiene que $D_{\hat{u}} f(\hat{x})$ existe para todo $\hat{x} \in U$, entonces $D_{\hat{u}} f$ es nuevamente una función de $U \subset \mathbb{R}^n$ en $\mathbb{R}$, es decir, $D_{\hat{u}} f : U \subset \mathbb{R}^n \to \mathbb{R}$. De esta forma, podemos trabajar todos los conceptos de derivación que hasta ahora hemos definido para esta nueva función. En particular, podemos ahora preguntarnos por la existencia de la derivada direccional de $D_{\hat{u}} f$ en un punto $\hat{x} \in U$, en la dirección de cualquier otro vector unitario $\hat{v} \in \mathbb{R}^n$. De acuerdo con la definición 4.1, dicha derivada se denotaría por $D_{\hat{v}}(D_{\hat{u}} f)(\hat{x})$ y estaría dada por

$$D_{\hat{v}}(D_{\hat{u}} f)(\hat{x}) := \lim_{h \to 0} \frac{D_{\hat{u}} f(\hat{x} + h\hat{v}) - D_{\hat{u}} f(\hat{x})}{h}.$$

Más aún, si ahora para $\hat{u}$ y $\hat{v}$ fijos, se tiene que $D_{\hat{v}}(D_{\hat{u}}f)(\hat{x})$ existe para toda $\hat{x} \in U$, entonces $D_{\hat{v}}(D_{\hat{u}}f)$ será otra vez una función de $U \subset \mathbb{R}^n$ en $\mathbb{R}$ y por tanto tiene sentido tomar otro vector unitario $\hat{w} \in \mathbb{R}^n$ y preguntarnos por la existencia de la derivada direccional de $D_{\hat{v}}(D_{\hat{u}}f)$ en un punto $\hat{x} \in U$, en la dirección del vector $\hat{w}$. Es decir, si existe

$$D_{\hat{w}}(D_{\hat{v}}(D_{\hat{u}}f))(\hat{x}) := \lim_{h\to 0} \frac{D_{\hat{v}}(D_{\hat{u}}f)(\hat{x}+h\hat{w}) - D_{\hat{v}}(D_{\hat{u}}f)(\hat{x})}{h}.$$

Antes de seguir por este camino, tomemos un respiro y notemos lo siguiente: seguramente a estas alturas, tomando en cuenta resultados como los de la proposición 4.13, el lector ya intuye la importancia de las derivadas direccionales de una función f en la dirección de los vectores de una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$, es decir, de las derivadas parciales.

Por esta razón, y aun cuando las ideas planteadas en los párrafos anteriores se pueden seguir trabajando en ese contexto más general, nos concentraremos en el caso particular de las derivadas parciales, ideas que dejaremos expresadas en la siguiente

Definición 4.9 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $r > 0$ tal que $B_r(\hat{x}_0) \subset U$, y $x_1, \ldots, x_n$ las coordenadas determinadas por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$ de $\mathbb{R}^n$. Si $\frac{\partial f}{\partial x_i}(\hat{x})$ existe para toda $\hat{x} \in B_r(\hat{x}_0)$, definimos la segunda derivada parcial de f con respecto a x_i y respecto a x_j en $\hat{x}_0$, que denotaremos por $\frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0)$, como la derivada parcial de la función $\frac{\partial f}{\partial x_i}$ con respecto de x_j en $\hat{x}_0$. Es decir,*

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) := \lim_{h\to 0} \frac{\frac{\partial f}{\partial x_i}(\hat{x}_0 + h\hat{e}_j) - \frac{\partial f}{\partial x_i}(\hat{x}_0)}{h}.$$

Si $j = i$, denotamos la derivada anterior por $\frac{\partial^2 f}{\partial x_i^2}(\hat{x}_0)$ y decimos que es la segunda derivada parcial de f con respecto de x_i, en $\hat{x}_0$.

Como en el caso de la (primera) derivada parcial, desde un punto de vista práctico, el cálculo de las segundas derivadas parciales se realiza siguiendo el mismo esquema: en cada paso, derivar sólo con respecto a una variable,

suponiendo que las demás son constantes. Como es de esperarse, lo que ahora sigue es un ejemplo que ilustre este concepto.

Ejemplo 4.13 *Sea* $f : \mathbb{R}^2 \to \mathbb{R}$ *dada por la expresión*

$$f(x, y) = x \cos(xy) + y \operatorname{sen}(xy).$$

Lo que haremos será calcular todas las posibles segundas derivadas parciales de f *(que son cuatro) para cualquier* $\hat{x} = (x, y) \in \mathbb{R}^2$.
Para ello, notemos primero que

$$\frac{\partial f}{\partial x}(x, y) = -xy \operatorname{sen}(xy) + \cos(xy) + y^2 \cos(xy)$$

y que

$$\frac{\partial f}{\partial y}(x, y) = -x^2 \operatorname{sen}(xy) + xy \cos(xy) + \operatorname{sen}(xy).$$

Por tanto, se tiene que

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2}(x, y) &= \frac{\partial \left(\frac{\partial f}{\partial x}\right)}{\partial x}(x, y) \\ &= -xy^2 \cos(xy) - y \operatorname{sen}(xy) - y \operatorname{sen}(xy) - y^3 \operatorname{sen}(xy) \\ &= -xy^2 \cos(xy) - 2y \operatorname{sen}(xy) - y^3 \operatorname{sen}(xy)\end{aligned}$$

y que

$$\begin{aligned}&\frac{\partial^2 f}{\partial y \partial x}(x, y) \\ &= \frac{\partial \left(\frac{\partial f}{\partial x}\right)}{\partial y}(x, y) \\ &= -x^2 y \cos(xy) - x \operatorname{sen}(xy) - x \operatorname{sen}(xy) - xy^2 \operatorname{sen}(xy) + 2y \cos(xy) \\ &= -x^2 y \cos(xy) - 2x \operatorname{sen}(xy) - xy^2 \operatorname{sen}(xy) + 2y \cos(xy).\end{aligned}$$

Finalmente,

$$\begin{aligned}
&\frac{\partial^2 f}{\partial x \partial y}(x,y)\\
&= \frac{\partial\left(\frac{\partial f}{\partial y}\right)}{\partial x}(x,y)\\
&= -x^2 y\cos(xy) - 2x\,\mathrm{sen}(xy) - xy^2\,\mathrm{sen}(xy) + y\cos(xy) + y\cos(xy)\\
&= -x^2 y\cos(xy) - 2x\,\mathrm{sen}(xy) - xy^2\,\mathrm{sen}(xy) + 2y\cos(xy)
\end{aligned}$$

y

$$\begin{aligned}
\frac{\partial^2 f}{\partial y^2}(x,y) &= \frac{\partial\left(\frac{\partial f}{\partial y}\right)}{\partial y}(x,y)\\
&= -x^3\,\mathrm{sen}(xy) - x^2 y\,\mathrm{sen}(xy) + x\cos(xy) + x\cos(xy)\\
&= -x^3\,\mathrm{sen}(xy) - x^2 y\,\mathrm{sen}(xy) + 2x\cos(xy).
\end{aligned}$$

Del ejemplo anterior es muy importante hacer dos observaciones con respecto al cálculo de las segundas derivadas parciales en las que aparecen ambas variables (x y y), y que se les conoce con el nombre de *derivadas parciales mixtas* (o *cruzadas*).

Una, que se refiere al orden en que se toman estas variables para calcular las correspondientes derivadas, y que es de derecha a izquierda, según aparezcan en el denominador correspondiente; y dos, que en este ejemplo, estas segundas derivadas parciales resultaron ser iguales.

De esta segunda observación, sin duda surge la pregunta de si dicha igualdad siempre se cumple. La respuesta es que no. En el problema 27 el lector encontrará un ejemplo de que esta identidad no siempre sucede. Por nuestra parte, lo siguiente que haremos será justo enunciar un resultado que nos proporciona las condiciones suficientes para que dicha igualdad sí se cumpla.

Un punto importante en la prueba de este resultado tiene que ver con un hecho muy sencillo, relacionado con los valores de una función sobre cualesquiera cuatro puntos de su dominio (que los podemos pensar como cuatro

vértices de un cuadrilátero), y el cálculo de ciertas diferencias tomadas a partir de estos valores.

Nótese que, si $\hat{x}_1, \hat{x}_2, \hat{x}_3$ y $\hat{x}_4$ son puntos del dominio de una función f, entonces se tiene que

$$(f(\hat{x}_3) - f(\hat{x}_2)) - (f(\hat{x}_4) - f(\hat{x}_1)) = (f(\hat{x}_3) - f(\hat{x}_4)) - (f(\hat{x}_2) - f(\hat{x}_1)).$$

Teorema 4.1 (de las parciales cruzadas) *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $i, j \in \{1, \ldots, n\}$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Si las segundas derivadas parciales $\frac{\partial^2 f}{\partial x_j \partial x_i}$ y $\frac{\partial^2 f}{\partial x_i \partial x_j}$ existen para cada $\hat{x} \in B_r(\hat{x}_0)$ y son continuas en $\hat{x}_0$, entonces*

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0).$$

Demostración. Sea $B_r(\hat{0}) = \{(h_1, h_2) \in \mathbb{R}^2 \mid \|(h_1, h_2)\| < r\}$. Nótese que si $(h_1, h_2) \in B_r(\hat{0})$, entonces $\hat{x}_0 + h_1\hat{e}_i + h_2\hat{e}_j$, $\hat{x}_0 + h_1\hat{e}_i$ y $\hat{x}_0 + h_2\hat{e}_j$ pertenecen a $B_r(\hat{x}_0) \subset \mathbb{R}^n$, de modo que podemos definir $H : B_r(\hat{0}) \subset \mathbb{R}^2 \to \mathbb{R}$ como

$$\begin{aligned} &H(h_1, h_2) \\ &= (f(\hat{x}_0 + h_1\hat{e}_i + h_2\hat{e}_j) - f(\hat{x}_0 + h_1\hat{e}_i)) - (f(\hat{x}_0 + h_2\hat{e}_j) - f(\hat{x}_0)). \end{aligned} \quad (4.35)$$

Ahora, dado $(h_1, h_2) \in B_r(\hat{0})$, definimos

$$g : \left\{ x \in \mathbb{R} \mid |x| < \sqrt{r^2 - h_2^2} \right\} = \left(-\sqrt{r^2 - h_2^2}, \sqrt{r^2 - h_2^2} \right) \subset \mathbb{R} \to \mathbb{R}$$

como

$$g(x) = f(\hat{x}_0 + x\hat{e}_i + h_2\hat{e}_j) - f(\hat{x}_0 + x\hat{e}_i).$$

Ya que $h_1 \in \left(-\sqrt{r^2 - h_2^2}, \sqrt{r^2 - h_2^2} \right)$, se tiene que

$$\begin{aligned} H(h_1, h_2) &= (f(\hat{x}_0 + h_1\hat{e}_i + h_2\hat{e}_j) - f(\hat{x}_0 + h_1\hat{e}_i)) - (f(\hat{x}_0 + h_2\hat{e}_j) - f(\hat{x}_0)) \\ &= g(h_1) - g(0). \end{aligned}$$

Por otra parte, observe que por el problema 2 de este capítulo, g es derivable en su dominio y

$$g'(x) = \frac{\partial f}{\partial x_i}(\hat{x}_0 + x\hat{e}_i + h_2\hat{e}_j) - \frac{\partial f}{\partial x_i}(\hat{x}_0 + x\hat{e}_i),$$

de modo que por el Teorema de Valor Medio (para funciones de $\mathbb{R}$ en $\mathbb{R}$), existe

$$\xi \in \left(-\sqrt{r^2 - h_2^2}, \sqrt{r^2 - h_2^2}\right)$$

tal que

$$\begin{aligned} H(h_1, h_2) &= g(h_1) - g(0) \\ &= g'(\xi)h_1 \\ &= \left(\frac{\partial f}{\partial x_i}(\hat{x}_0 + \xi\hat{e}_i + h_2\hat{e}_j) - \frac{\partial f}{\partial x_i}(\hat{x}_0 + \xi\hat{e}_i)\right) h_1. \end{aligned}$$

Finalmente, como

$$[\hat{x}_0 + \xi\hat{e}_i, \hat{x}_0 + \xi\hat{e}_i + h_2\hat{e}_j] \subset B_r(\hat{x}_0)$$

y

$$(\hat{x}_0 + \xi\hat{e}_i + h_2\hat{e}_j) - (\hat{x}_0 + \xi\hat{e}_i) = h_2\hat{e}_j,$$

por la proposición 4.8 (aplicada a la función $\frac{\partial f}{\partial x_i}$) se tiene que existe $\eta \in (0, |h_2|)$ tal que

$$\begin{aligned} H(h_1, h_2) &= \left(\frac{\partial f}{\partial x_i}(\hat{x}_0 + \xi\hat{e}_i + h_2\hat{e}_j) - \frac{\partial f}{\partial x_i}(\hat{x}_0 + \xi\hat{e}_i)\right) h_1 \\ &= \frac{\partial \left(\frac{\partial f}{\partial x_i}\right)}{\partial x_j}(\hat{x}_0 + \xi\hat{e}_i + \eta\hat{e}_j)h_1h_2 \\ &= \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0 + \xi\hat{e}_i + \eta\hat{e}_j)h_1h_2. \end{aligned}$$

Por lo tanto, si $h_1h_2 \neq 0$, por la continuidad de $\frac{\partial^2 f}{\partial x_j \partial x_i}$ en $\hat{x}_0$, y del hecho de que $(\xi, \eta) \to (0,0)$ si $(h_1, h_2) \to (0,0)$, concluimos que

$$\begin{aligned}\lim_{(h_1,h_2)\to(0,0)} \frac{H(h_1,h_2)}{h_1h_2} &= \lim_{(\xi,\eta)\to(0,0)} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0 + \xi\hat{e}_i + \eta\hat{e}_j) \\ &= \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0).\end{aligned}$$

Observe ahora que para cada $(h_1, h_2) \in B_r(\hat{0})$, la función H definida en 4.35 también se puede escribir como

$$H(h_1,h_2) = (f(\hat{x}_0 + h_1\hat{e}_i + h_2\hat{e}_j) - f(\hat{x}_0 + h_2\hat{e}_j)) - (f(\hat{x}_0 + h_1\hat{e}_i) - f(\hat{x}_0))$$

de tal forma que, dado $(h_1, h_2) \in B_r(\hat{0})$, ahora definimos

$$\tilde{g} : \left\{x \in \mathbb{R} \mid |x| < \sqrt{r^2 - h_1^2}\right\} = \left(-\sqrt{r^2 - h_1^2}, \sqrt{r^2 - h_1^2}\right) \subset \mathbb{R} \to \mathbb{R}$$

como

$$\tilde{g}(x) = f(\hat{x}_0 + h_1\hat{e}_i + x\hat{e}_j) - f(\hat{x}_0 + x\hat{e}_j).$$

Ya que $h_2 \in \left(-\sqrt{r^2 - h_1^2}, \sqrt{r^2 - h_1^2}\right)$, se tiene que

$$H(h_1,h_2) = \tilde{g}(h_2) - \tilde{g}(0).$$

Procediendo con la función $\tilde{g}$ como lo hicimos anteriormente con la función g, se tiene que existe $\xi' \in \left(-\sqrt{r^2 - h_1^2}, \sqrt{r^2 - h_1^2}\right)$ tal que

$$\begin{aligned}H(h_1,h_2) &= \tilde{g}(h_2) - \tilde{g}(0) \\ &= \tilde{g}'(\xi')h_2 \\ &= \left(\frac{\partial f}{\partial x_j}(\hat{x}_0 + h_1\hat{e}_i + \xi'\hat{e}_j) - \frac{\partial f}{\partial x_j}(\hat{x}_0 + \xi'\hat{e}_j)\right) h_2.\end{aligned}$$

Como

$$\left[\hat{x}_0 + \xi'\hat{e}_j, \hat{x}_0 + h_1\hat{e}_i + \xi'\hat{e}_j\right] \subset B_r(\hat{x}_0),$$

con

$$\left(\hat{x}_0 + h_1\hat{e}_i + \xi'\hat{e}_j\right) - \left(\hat{x}_0 + \xi'\hat{e}_j\right) = h_1\hat{e}_i,$$

nuevamente por la proposición 4.8 (aplicada a la función $\frac{\partial f}{\partial x_j}$) se tiene que existe $\eta' \in (0, |h_1|)$ tal que

$$\begin{aligned} H(h_1, h_2) &= \left(\frac{\partial f}{\partial x_j}(\hat{x}_0 + h_1\hat{e}_i + \xi'\hat{e}_j) - \frac{\partial f}{\partial x_j}(\hat{x}_0 + \xi'\hat{e}_j)\right) h_2 \\ &= \frac{\partial\left(\frac{\partial f}{\partial x_j}\right)}{\partial x_i}(\hat{x}_0 + \eta'\hat{e}_i + \xi'\hat{e}_j) h_1 h_2 \\ &= \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0 + \eta'\hat{e}_i + \xi'\hat{e}_j) h_1 h_2. \end{aligned}$$

De esta forma, si $h_1 h_2 \neq 0$, por la continuidad de $\frac{\partial^2 f}{\partial x_i \partial x_j}$ en $\hat{x}_0$ y que $(\xi', \eta') \to (0,0)$ si $(h_1, h_2) \to (0,0)$, concluimos que

$$\begin{aligned} \lim_{(h_1,h_2)\to(0,0)} \frac{H(h_1, h_2)}{h_1 h_2} &= \lim_{(\xi',\eta')\to(0,0)} \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0 + \eta'\hat{e}_i + \xi'\hat{e}_j) \\ &= \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0). \end{aligned}$$

Por tanto, como el límite de cualquier función, si existe, es único, tenemos que

$$\begin{aligned} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) &= \lim_{(h_1,h_2)\to(0,0)} \frac{H(h_1, h_2)}{h_1 h_2} \\ &= \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0), \end{aligned}$$

que es la identidad que deseabamos probar. ■

Como mencionamos al inicio de esta sección, podemos calcular sucesivamente tantas derivadas direccionales (en la misma o en diferentes direcciones) como nos lo permita la función con la que estemos trabajando.

Dado que en términos prácticos es suficiente hacer esto para los vectores de una base ortonormal de $\mathbb{R}^n$, lo siguiente que haremos será "extender" la definición 4.9 para cualquier $k \in \mathbb{N}$.

Definición 4.10 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $r > 0$ tal que $B_r(\hat{x}_0) \subset U$, $k \in \mathbb{N}$ y $x_1, \ldots, x_n$ las coordenadas determinadas por una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$ de $\mathbb{R}^n$. Si $i_1, \ldots, i_k, i_{k+1} \in \{1, \ldots, n\}$ son tales que $\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}(\hat{x})$ existe para toda $\hat{x} \in B_r(\hat{x}_0)$, definimos la $k+1$ derivada parcial de f con respecto a $x_{i_1}, \ldots, x_{i_k}, x_{i_{k+1}}$, en $\hat{x}_0$, que denotaremos por $\frac{\partial^{k+1} f}{\partial x_{i_{k+1}} \partial x_{i_k} \cdots \partial x_{i_1}}(\hat{x}_0)$, como la derivada parcial con respecto de $x_{i_{k+1}}$ de la función $\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}$ en $\hat{x}_0$. Es decir,*

$$\frac{\partial^{k+1} f}{\partial x_{i_{k+1}} \partial x_{i_k} \cdots \partial x_{i_1}}(\hat{x}_0) := \lim_{h \to 0} \frac{\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}(\hat{x}_0 + h\hat{e}_{i_{k+1}}) - \frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}(\hat{x}_0)}{h}.$$

Si $i_1 = i_2 = \cdots = i_k = i_{k+1} = i$, la derivada anterior la denotamos por $\frac{\partial^{k+1} f}{\partial x_i^{k+1}}(\hat{x}_0)$ y decimos que es la $k+1$ derivada parcial de f con respecto de x_i, en $\hat{x}_0$.

A propósito de la notación usada en esta definición, y en concordancia con la última parte, en general, si en la expresión $\partial x_{i_k} \cdots \partial x_{i_1}$ existen m índices consecutivos iguales, esta parte de la expresión la sustituiremos por ∂x_i^m, en donde x_i es la variable que se repite m veces consecutivas.

Por ejemplo, la derivada parcial de orden 6

$$\frac{\partial^6 f}{\partial x_3 \partial x_2 \partial x_2 \partial x_3 \partial x_1 \partial x_1}(\hat{x}_0)$$

se podrá escribir como

$$\frac{\partial^6 f}{\partial x_3 \partial x_2^2 \partial x_3 \partial x_1^2}(\hat{x}_0)$$

Dada $k \in \mathbb{N}$, como seguramente el lector recordará de sus cursos de álgebra, existen n^k formas de elegir ordenaciones (con repetición) con elementos

del conjunto $\{1, \ldots, n\}$, de tal forma que existen n^k derivadas parciales de orden k para una función f de $\mathbb{R}^n$ en $\mathbb{R}$. Cuando todas estas derivadas parciales existen en los puntos de un conjunto (abierto) sobre el cual esté definida la función f, y además son continuas en este mismo conjunto, diremos que f *es una función de clase* C^k. Este concepto jugará un papel muy importante más adelante, y lo dejaremos plasmado en la siguiente

Definición 4.11 *Sean* $f : U \subset \mathbb{R}^n \to \mathbb{R}$ *y* $k \in \mathbb{N}$. *Decimos que* f *es una función de clase* C^k *en* U *si existen todas las derivadas parciales de orden* k *de* f *en cada punto de* U, *y además estas derivadas parciales son continuas en cada punto de* U.

El concepto anterior está muy relacionado con las hipótesis de la proposición 4.13 y el teorema 4.1, razón por la cual enunciaremos el siguiente par de proposiciones y dejaremos su prueba al lector.

Proposición 4.17 *Si* $f : U \subset \mathbb{R}^n \to \mathbb{R}$ *es de clase* C^1 *en* U, *entonces* f *es derivable para toda* $\hat{x} \in U$.

Proposición 4.18 *Si* $f : U \subset \mathbb{R}^n \to \mathbb{R}$ *es de clase* C^2 *en* U, *entonces*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x})$$

para toda $\hat{x} \in U$ *y para todas* $i, j \in \{1, \ldots, n\}$.

Cuando una función es de clase C^k, muchas de sus derivadas parciales de orden k coinciden. En efecto, si en dos derivadas parciales de orden k de una función f se deriva con respecto a las mismas variables, y además se deriva (con respecto de cada una de ellas y sin importar el orden) el mismo número de veces, entonces estas derivadas parciales son iguales. Enunciar con toda precisión (¡y probar!) la afirmación anterior no está dentro de los objetivos de este texto, y sólo baste decir que en el fondo su prueba se basa en el teorema 4.1.

Antes de iniciar con otro tema, es importante enfatizar qué conceptos hemos definido (¡y cuáles no hemos definido!) en esta sección.

Como lo dice su nombre, aquí sólo definimos el concepto de derivadas direccionales de orden superior de una función f, lo que no significa que hayamos definido el concepto de derivada (global) de orden superior de f.

Como hemos venido insistiendo, la derivada de una función f es una función que a cada $\hat{x}$ del dominio de f le asocia una función lineal de $\mathbb{R}^n$ en $\mathbb{R}$ que denotamos por $Df(\hat{x})$. Siguiendo la idea de que la segunda derivada de una función f debería de ser la derivada de la derivada de f, para definir este concepto tendríamos que definir lo que significa derivar una función cuyo dominio está contenido en $\mathbb{R}^n$ y que tiene como contradominio al conjunto de las funciones lineales de $\mathbb{R}^n$ en $\mathbb{R}$.

Seguramente el lector estará de acuerdo en que, abordar el problema que se planteó en el párrafo precedente, escapa a los objetivos del texto. No obstante lo anterior, con los conceptos desarrollados en esta sección nos es suficiente para abordar un tema muy importante relacionado con la "aproximación" de una función cerca de un punto, tema que abordaremos en la siguiente sección.

4.5. Aproximación polinomial

Como el lector recordará, el concepto de derivada (global) de una función f de $\mathbb{R}^n$ en $\mathbb{R}$ en un punto $\hat{x}_0$, está íntimamente relacionado con el problema de aproximar a dicha función alrededor (o cerca) de $\hat{x}_0$ por medio de una función lineal. En términos más concretos, la derivabilidad de f en $\hat{x}_0$ se reduce al hecho de que la función

$$P(\hat{x}) = Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0)$$

es la única función "de este tipo" que tiene la propiedad de que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - P(\hat{x})}{\|\hat{x} - \hat{x}_0\|} = 0. \tag{4.36}$$

Si ahora recordamos que

$$Df(\hat{x}_0)(\hat{x} - \hat{x}_0) = \nabla f(\hat{x}_0) \cdot (\hat{x} - \hat{x}_0),$$

tendremos que en un sistema de coordenadas $x_1, \ldots, x_n$, en donde $\hat{x}_0 = \left(x_1^{(0)}, \ldots, x_n^{(0)}\right)$, la función P toma la forma

$$\begin{aligned} P(\hat{x}) &= P(x_1, \ldots, x_n) \\ &= \nabla f(\hat{x}_0) \cdot (\hat{x} - \hat{x}_0) + f(\hat{x}_0) \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0) \left(x_i - x_i^{(0)}\right) + f(\hat{x}_0), \end{aligned} \tag{4.37}$$

de donde concluimos que P es una función *polinomial* de grado a lo más 1 de las coodenadas (o variables) $x_1, \ldots, x_n$, con la particularidad de que $P(\hat{x}_0) = f(\hat{x}_0)$.

Visto de esta forma, con base en la proposición 4.13 podemos asegurar que si las derivadas parciales $\frac{\partial f}{\partial x_i}$ existen en una vecindad del punto $\hat{x}_0$ y son continuas en éste (lo que se satisface sobradamente si f es de clase C^1 en una vecindad de $\hat{x}_0$), entonces el polinomio de grado a lo más 1 en las variables $x_1, \ldots, x_n$ dado por 4.37 es el único polinomio que satisface la condición 4.36.

Como seguramente el lector ya se estará imaginando, la siguiente pregunta que nos haremos será la siguiente: ¿qué propiedades deberá tener una función f para que podamos asegurar que existe una función polinomial de grado a lo más 2 en las variables $x_1, \ldots, x_n$, con la particularidad de que valga lo mismo que f en $\hat{x}_0$, tal que *se parezca mucho a f alrededor de $\hat{x}_0$*?

Para abordar este problema, empezaremos analizando el caso de una función definida en $\mathbb{R}^2$. Supongamos entonces que $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ y que $\hat{x}_0 = (x_0, y_0) \in U$.

Primero notemos que una función $P_2(x, y)$ polinomial en las variables x y y de grado a lo más 2, es de la forma

$$P_2(x, y) = Ax^2 + By^2 + Cxy + Dx + Ey + F.$$

Si se desea que esta función tome el valor $f(\hat{x}_0)$ cuando $(x, y) = (x_0, y_0)$, entonces deberá ser de la forma

$$\begin{aligned} P_2(x, y) = A(x - x_0)^2 + B(y - y_0)^2 + C(x - x_0)(y - y_0) \\ + D(x - x_0) + E(y - y_0) + f(\hat{x}_0), \end{aligned} \tag{4.38}$$

por lo que nuestro problema se "reduce" a encontrar los coeficientes A, B, C, D y E.

Ahora, si el criterio para decir que P_2 se parece mucho a f alrededor de $\hat{x}_0$ es que satisfaga la condición 4.36, dado que

$$(x-x_0)^2, (y-y_0)^2, |(x-x_0)(y-y_0)| \leq \|\hat{x}-\hat{x}_0\|^2,$$

entonces

$$\begin{aligned}\lim_{(x,y)\to(x_0,y_0)} \frac{(x-x_0)^2}{\|\hat{x}-\hat{x}_0\|} &= \lim_{(x,y)\to(x_0,y_0)} \frac{(y-y_0)^2}{\|\hat{x}-\hat{x}_0\|} \\ &= \lim_{(x,y)\to(x_0,y_0)} \frac{(x-x_0)(y-y_0)}{\|\hat{x}-\hat{x}_0\|} \\ &= 0.\end{aligned}$$

Por lo tanto,

$$\begin{aligned}0 &= \lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x})-P_2(\hat{x})}{\|\hat{x}-\hat{x}_0\|} \\ &= \lim_{\hat{x}\to\hat{x}_0} \left(\frac{f(\hat{x}) - \big(A(x-x_0)^2 + B(y-y_0)^2\big)}{\|\hat{x}-\hat{x}_0\|} \right. \\ &\quad \left. + \frac{C(x-x_0)(y-y_0) + D(x-x_0) + E(y-y_0) + f(\hat{x}_0)}{\|\hat{x}-\hat{x}_0\|} \right) \\ &= \lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - (D(x-x_0) + E(y-y_0) + f(\hat{x}_0))}{\|\hat{x}-\hat{x}_0\|},\end{aligned}$$

lo cual es equivalente a que f sea derivable en el punto $\hat{x}_0$. De esta forma, se deberá tener que

$$D = \frac{\partial f}{\partial x}(\hat{x}_0) \qquad \text{y} \qquad E = \frac{\partial f}{\partial y}(\hat{x}_0).$$

En conclusión, si f es derivable en $\hat{x}_0$, entonces la función polinomial

$$\begin{aligned}P_2(x,y) = A(x-x_0)^2 + B(y-y_0)^2 + C(x-x_0)(y-y_0) \\ + \frac{\partial f}{\partial x}(\hat{x}_0)(x-x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)(y-y_0) + f(\hat{x}_0) \quad (4.39)\end{aligned}$$

cumple la condición 4.36 para cualesquiera A, B y C números reales.

Tal vez esta última parte no sea del todo una buena noticia, pues nos asegura que hay una gran cantidad de funciones polinomiales de grado 2 que se "parecen" mucho a f alrededor de $\hat{x}_0$, lo que sin duda nos deja con la pregunta de si de entre todas ellas no habrá alguna que sea "la mejor" de todas.

Lo que haremos a continuación será mostrar que sí existe un criterio con base en el cual podemos elegir, de entre todas las funciones polinomiales dadas por 4.39, la que "se parece más" a f alrededor de $\hat{x}_0$. Para ello, basta con hacer notar que una forma de interpretar la multicitada condición 4.36 es que el número $f(\hat{x}) - P(\hat{x})$ se hace 0 mucho más "rápido" que el número $\|\hat{x} - \hat{x}_0\|$.

Con base en esta interpretación, y dado que, si $\|\hat{x} - \hat{x}_0\| \leq 1$, se tiene que $\|\hat{x} - \hat{x}_0\|^2 \leq \|\hat{x} - \hat{x}_0\|$, entonces pedir que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - P(\hat{x})}{\|\hat{x} - \hat{x}_0\|^2} = 0 \tag{4.40}$$

sin duda es una condición "más fuerte" que la condición 4.36 (de hecho, por el problema 37 del capítulo 2, sabemos que si se cumple la condición anterior, entonces se cumple la condición 4.36 (¡pero no lo recíproco!)). Por esta razón, si suponemos que una función polinomial de grado a lo más 2 satisface la condición 4.40, podemos dar por hecho que ya es de la forma 4.39.

Supongamos entonces que P_2 es una función polinomial como en 4.39 que satisface la condición 4.40. Es decir, supongamos que

$$\begin{aligned}
&\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - P_2(\hat{x})}{\|\hat{x} - \hat{x}_0\|^2} \\
&= \lim_{\hat{x}\to\hat{x}_0} \left(\frac{f(\hat{x}) - \big(A(x-x_0)^2 + B(y-y_0)^2 + C(x-x_0)(y-y_0)\big)}{\|\hat{x} - \hat{x}_0\|^2} \right. \\
&\qquad \left. + \frac{\frac{\partial f}{\partial x}(\hat{x}_0)(x-x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)(y-y_0) + f(\hat{x}_0)}{\|\hat{x} - \hat{x}_0\|^2} \right) \\
&= 0.
\end{aligned}$$

Si nos aproximamos a (x_0, y_0) por puntos de la forma $(x, y) = (x_0+h, y_0)$, se deberá tener que

$$\begin{aligned} 0 &= \lim_{h\to 0} \frac{f(x_0+h, y_0) - \left(Ah^2 + \frac{\partial f}{\partial x}(x_0, y_0)h + f(x_0, y_0)\right)}{h^2} \\ &= \lim_{h\to 0} \left[\frac{f(x_0+h, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - f(x_0, y_0)}{h^2} - A\right], \end{aligned}$$

de modo que al cociente

$$\frac{f(x_0+h, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - f(x_0, y_0)}{h^2}$$

lo podemos tratar como un cociente de dos funciones que dependen de la variable (real) h.

Ahora, si suponemos que f y sus derivadas parciales son funciones derivables en todos los puntos de una vecindad del punto $\hat{x}_0$ (lo cual se puede garantizar suponiendo que f es de clase C^2 en una vecindad de este punto), aplicando (las veces que haga falta) la regla de L'Hospital para este tipo de funciones (y por supuesto la regla de la caden probada en la proposición 4.16), tenemos que

$$\begin{aligned} &\lim_{h\to 0} \frac{f(x_0+h, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)h - f(x_0, y_0)}{h^2} \\ &\quad = \lim_{h\to 0} \frac{\frac{\partial f}{\partial x}(x_0+h, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)}{2h} \\ &\quad = \frac{1}{2}\lim_{h\to 0} \frac{\frac{\partial f}{\partial x}(x_0+h, y_0) - \frac{\partial f}{\partial x}(x_0, y_0)}{h} \\ &\quad = \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x_0, y_0), \end{aligned}$$

de donde concluimos que

$$A = \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x_0, y_0).$$

Aproximándonos a (x_0, y_0) por puntos de la forma $(x, y) = (x_0, y_0 + h)$, por un procedimiento análogo se concluye que

$$B = \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(x_0, y_0).$$

Finalmente, para deducir el valor del coeficiente C, nos aproximaremos a (x_0, y_0) por puntos de la forma $(x, y) = (x_0 + h, y_0 + h)$.

En este caso, usando nuevamente la regla de L'Hospital para obtener la segunda y tercera identidad[2], se deberá tener que

$$\begin{aligned}
0 &= \lim_{h\to 0}\left[\frac{f(x_0+h, y_0+h) - \left(\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(\hat{x}_0)h^2 + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(\hat{x}_0)h^2\right)}{h^2}\right. \\
&\qquad \left. - \frac{Ch^2 + \frac{\partial f}{\partial x}(\hat{x}_0)h + \frac{\partial f}{\partial y}(\hat{x}_0)h + f(\hat{x}_0)}{h^2}\right] \\
&= \lim_{h\to 0}\left[\frac{\frac{\partial f}{\partial x}(x_0+h, y_0+h) + \frac{\partial f}{\partial y}(x_0+h, y_0+h)}{2h}\right. \\
&\qquad \left. - \frac{\frac{\partial^2 f}{\partial x^2}(\hat{x}_0)h + \frac{\partial^2 f}{\partial y^2}(\hat{x}_0)h + 2Ch + \frac{\partial f}{\partial x}(\hat{x}_0) + \frac{\partial f}{\partial y}(\hat{x}_0)}{2h}\right] \\
&= \lim_{h\to 0}\frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(x_0+h, y_0+h) + \frac{\partial^2 f}{\partial y \partial x}(x_0+h, y_0+h)\right. \\
&\qquad + \frac{\partial^2 f}{\partial x \partial y}(x_0+h, y_0+h) + \frac{\partial^2 f}{\partial y^2}(x_0+h, y_0+h) \\
&\qquad \left. - \left(\frac{\partial^2 f}{\partial x^2}(\hat{x}_0) + \frac{\partial^2 f}{\partial y^2}(\hat{x}_0) + 2C\right)\right] \\
&= \frac{1}{2}\left(\frac{\partial^2 f}{\partial y \partial x}(\hat{x}_0) + \frac{\partial^2 f}{\partial x \partial y}(\hat{x}_0) - 2C\right),
\end{aligned}$$

[2]Es importante hacer notar que, por razones de espacio, las expresiones a las cuales se aplica la regla de L'Hospital tuvieron que escribirse como la diferencia de dos cocientes. Sin embargo, para la aplicación de dicha regla, estas expresiones tienen que considerarse como un solo cociente.

de donde obtenemos que

$$C = \frac{1}{2}\left(\frac{\partial^2 f}{\partial y \partial x}(\hat{x}_0) + \frac{\partial^2 f}{\partial x \partial y}(\hat{x}_0)\right).$$

De lo anterior concluimos que, si existen las segundas derivadas parciales de f (en al menos una vecindad del punto $\hat{x}_0$ y en este punto son continuas), y P_2 es una función polinomial de grado menor o igual a 2 que satisface la condición 4.40, entonces necesariamente P_2 está dado por

$$\begin{aligned} P_2(x,y) = \frac{1}{2}\Bigg(&\frac{\partial^2 f}{\partial x^2}(\hat{x}_0)(x-x_0)^2 + \frac{\partial^2 f}{\partial y^2}(y-y_0)^2 + \\ &\left(\frac{\partial^2 f}{\partial y \partial x}(\hat{x}_0) + \frac{\partial^2 f}{\partial x \partial y}(\hat{x}_0)\right)(x-x_0)(y-y_0)\Bigg) \\ &+ \frac{\partial f}{\partial x}(\hat{x}_0)(x-x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)(y-y_0) + f(\hat{x}_0). \quad (4.41) \end{aligned}$$

Es importante enfatizar que el resultado anterior se obtuvo bajo el supuesto de que f es de clase C^2 en una vecindad del punto $\hat{x}_0$, y de que existe una función polinomial de grado menor o igual a 2 que satisface la condición 4.40. La buena noticia es que, bajo las mismas hipótesis sobre f, lo recíproco también es cierto, es decir la función polinomial dada por 4.41 cumple la condición 4.40.

Seguramente a esta alturas el lector ya estará empezando a vislumbrar el resultado más general hacia el cual estamos dirigiendo nuestros pasos. Todo parece indicar que, si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$ son tales que f es de clase C^N en una vecindad del punto $\hat{x}_0$ para alguna $N \in \mathbb{N}$, entonces la función polinomial de grado a lo más N dada por

$$\begin{aligned} P_N(\hat{x}) = f(\hat{x}_0) &+ \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0)\left(x_i - x_i^{(0)}\right) + \cdots \\ &+ \frac{1}{N!}\sum_{i_1,\ldots,i_N=1}^{n} \frac{\partial^N f}{\partial x_{i_1}\cdots\partial x_{i_N}}(\hat{x}_0)\left(x_{i_1} - x_{i_1}^{(0)}\right)\cdots\left(x_{i_N} - x_{i_N}^{(0)}\right) \end{aligned}$$

en donde $\hat{x} = (x_1, \ldots, x_n)$ y $\hat{x}_0 = \left(x_1^{(0)}, \ldots, x_n^{(0)}\right)$, debe ser tal que

$$\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - P_N(\hat{x})}{\|\hat{x} - \hat{x}_0\|^N} = 0$$

y además es la única con esta propiedad.

Este resultado es conocido como el *Teorema de Taylor*, y aquí lo formularemos un poco diferente a como lo acabamos de describir en el párrafo anterior. También es importante mencionar que la prueba de este teorema se basa en el Teorema de Taylor, en su versión para funciones de $\mathbb{R}$ en $\mathbb{R}$.

Teorema 4.2 (de Taylor) *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $r > 0$ y $N \in \mathbb{N}$ tales que f es de clase C^{N+1} en $B_r(\hat{x}_0) \subset U$. Definimos el polinomio de Taylor de grado N de la función f en $\hat{x}_0$, que denotamos por $P_{N,f,\hat{x}_0}(\hat{x})$, como*

$$P_{N,f,\hat{x}_0}(\hat{x}) = f(\hat{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0) \left(x_i - x_i^{(0)}\right) + \cdots$$
$$+ \frac{1}{N!} \sum_{i_1, \ldots, i_N = 1}^{n} \frac{\partial^N f}{\partial x_{i_N} \cdots \partial x_{i_1}}(\hat{x}_0) \left(x_{i_1} - x_{i_1}^{(0)}\right) \cdots \left(x_{i_N} - x_{i_N}^{(0)}\right)$$

(en donde $\hat{x} = (x_1, \ldots, x_n)$ y $\hat{x}_0 = \left(x_1^{(0)}, \ldots, x_n^{(0)}\right)$). Definimos el residuo de orden N de f en $\hat{x}_0$, que denotamos por $R_{N,f,\hat{x}_0}(\hat{x})$, como

$$R_{N,f,\hat{x}_0}(\hat{x}) = f(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})$$

para $\hat{x} \in U$. $P_{N,f,\hat{x}_0}$ y $R_{N,f,\hat{x}_0}$ satisfacen que:

1. *si $\hat{x} \in B_r(\hat{x}_0)$ y hacemos $\hat{u} = \hat{x} - \hat{x}_0$, entonces existe $\hat{\xi} \in [\hat{x}_0, \hat{x}]$ tal que*

$$R_{N,f,\hat{x}_0}(\hat{x}_0 + \hat{h})$$
$$= \frac{1}{(N+1)!} \sum_{i_1, \ldots, i_N, i_{N+1} = 1}^{n} \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) u_{i_1} \cdots u_{i_N} u_{i_{N+1}}$$

en donde $u_j = x_j - x_j^{(0)}$ para cada $j \in \{1, \ldots, n\}$

2.
$$\lim_{\hat{x}\to\hat{x}_0} \frac{R_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x}-\hat{x}_0\|^N} = \lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x}-\hat{x}_0\|^N} = 0$$

3. $P_{N,f,\hat{x}_0}$ *es el único polinomio de grado menor o igual a* N *que satisface la condición del inciso* 2.

Demostración. Para la prueba del inciso (1), tomamos $\hat{x} \in B_r(\hat{x}_0)$, $\hat{x} \neq \hat{x}_0$ (el caso $\hat{x} = \hat{x}_0$ es inmediato tomando $\hat{\xi} = \hat{x}_0$), y definimos

$$g : (-r/\|\hat{x}-\hat{x}_0\|, r/\|\hat{x}-\hat{x}_0\|) \subset \mathbb{R} \to \mathbb{R}$$

como $g(t) = f(\hat{x}_0 + t(\hat{x}-\hat{x}_0))$.

Por el problema 33 (tomando $\hat{h} = \hat{x}-\hat{x}_0$) sabemos que g es de clase C^{N+1} en su dominio de tal forma que, por el Teorema de Taylor para funciones de $\mathbb{R}$ en $\mathbb{R}$, tenemos que

$$\begin{aligned}
&R_{N,g,0}(1)\\
&= \frac{1}{(N+1)!} g^{(N+1)}(\xi)(1-0)\\
&= \frac{1}{(N+1)!} g^{(N+1)}(\xi)\\
&= \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi}\right) u_{i_1} \cdots u_{i_N} u_{i_{N+1}},
\end{aligned}$$

en donde $\hat{\xi} = \hat{x}_0 + \xi(\hat{x}-\hat{x}_0)$ para alguna $\xi \in (0,1)$, de modo que $\hat{\xi} \in [\hat{x}_0, \hat{x}]$.

Por otra parte, como

$$\begin{aligned}
R_{N,g,0}(1) &= g(1) - P_{N,g,0}(1)\\
&= f(\hat{x}) - \sum_{i=0}^{N} \frac{g^{(i)}(0)}{i!}(1-0)^i\\
&= f(\hat{x}) - (g(0) + g'(0) + \frac{1}{2!}g''(0) + \cdots + \frac{1}{N!}g^{(N)}(0)
\end{aligned}$$

$$
\begin{aligned}
&= f(\hat{x}) - \left(f(\hat{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0) \left(x_i - x_i^{(0)} \right) + \cdots \right. \\
&\qquad \left. + \frac{1}{N!} \sum_{i_1,\ldots,i_N=1}^{n} \frac{\partial^N f}{\partial x_{i_N} \cdots \partial x_{i_1}}(\hat{x}_0) \left(x_{i_1} - x_{i_1}^{(0)} \right) \cdots \left(x_{i_N} - x_{i_N}^{(0)} \right) \right) \\
&= f(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x}) \\
&= R_{N,f,\hat{x}_0}(\hat{x}),
\end{aligned}
$$

concluimos que

$$
\begin{aligned}
&R_{N,f,\hat{x}_0}(\hat{x}) \\
&= \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi} \right) u_{i_1} \cdots u_{i_N} u_{i_{N+1}}
\end{aligned}
$$

para alguna $\hat{\xi} \in [\hat{x}_0, \hat{x}]$.

Ahora, dado que

$$
|u_i| = \left| x_i - x_i^{(0)} \right| \leq \|\hat{x} - \hat{x}_0\|
$$

para cada $i \in \{1, \ldots, n\}$, entonces

$$
\begin{aligned}
&|R_{N,f,\hat{x}_0}(\hat{x})| \\
&= \left| \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi} \right) u_{i_1} \cdots u_{i_N} u_{i_{N+1}} \right| \\
&\leq \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \left| \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi} \right) \right| |u_{i_1}| \cdots |u_{i_N}| \left| u_{i_{N+1}} \right| \\
&\leq \frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \left| \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi} \right) \right| \|\hat{x} - \hat{x}_0\|^{N+1} \\
&= \frac{\|\hat{x} - \hat{x}_0\|^{N+1}}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \left| \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi} \right) \right|,
\end{aligned}
$$

de tal forma que

$$\frac{|R_{N,f,\hat{x}_0}(\hat{x})|}{\|\hat{x}-\hat{x}_0\|^N} \leq \|\hat{x}-\hat{x}_0\| \left(\frac{1}{(N+1)!} \sum_{i_1,\ldots,i_N,i_{N+1}=1}^{n} \left| \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi} \right) \right| \right).$$

Por tanto, dado que f es de clase C^{N+1} en $B_r(\hat{x}_0)$, se tiene que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} \left(\hat{\xi} \right) = \frac{\partial^{N+1} f}{\partial x_{i_{N+1}} \partial x_{i_N} \cdots \partial x_{i_1}} (\hat{x}_0)$$

para todas $i_1, , i_N, i_{N+1} \in \{1, \ldots, n\}$. Así concluimos que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{R_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x}-\hat{x}_0\|^N} = 0,$$

lo cual prueba el inciso (2).

Finalmente, para la prueba del inciso (3), supongamos que P_N es otro polinomio de grado menor o igual a N para el cual se satisface que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - P_N(\hat{x})}{\|\hat{x}-\hat{x}_0\|^N} = 0.$$

Dado que

$$\begin{aligned} \lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x}-\hat{x}_0\|^N} &= \lim_{\hat{x}\to\hat{x}_0} \frac{R_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x}-\hat{x}_0\|^N} \\ &= 0, \end{aligned}$$

restando estos dos últimos límites tenemos que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{P_N(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})}{\|\hat{x}-\hat{x}_0\|^N} = 0,$$

de tal forma que, si hacemos $P(\hat{x}) = P_N(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x})$, se tiene que P es un polinomio de grado menor o igual a N con la propiedad de que

$$\lim_{\hat{x}\to\hat{x}_0} \frac{P(\hat{x})}{\|\hat{x}-\hat{x}_0\|^N} = 0.$$

El siguiente paso será demostrar que un polinomio con estas características tiene que ser la función constante 0. Para ello, tomemos $\hat{x} \in \mathbb{R}^n$, fija. Dado que la identidad $P(\hat{x}_0) = 0$ se sigue inmediatamente del límite anterior y de la continuidad de P, supondremos que $\hat{x} \neq \hat{x}_0$.

Definimos $p(t) = P(\hat{x}_0 + t(\hat{x} - \hat{x}_0))$ para $t \in \mathbb{R}$; nótese que p será un polinomio en la variable t de grado a lo más N, con la propiedad de que

$$\begin{aligned}\lim_{t\to 0}\left|\frac{p(t)}{t^N}\right| &= \|\hat{x} - \hat{x}_0\|^N \lim_{t\to 0} \frac{|P(\hat{x}_0 + t(\hat{x} - \hat{x}_0))|}{\|(\hat{x}_0 + t(\hat{x} - \hat{x}_0)) - \hat{x}_0\|^N} \\ &= 0,\end{aligned}$$

o equivalentemente, que

$$\lim_{t\to 0} \frac{p(t)}{t^N} = 0.$$

Por lo tanto, como para el caso de una variable ya sabemos que un polinomio con estas características tiene que ser la constante cero, concluimos que

$$\begin{aligned}0 &= p(1) \\ &= P(\hat{x}) \\ &= P_N(\hat{x}) - P_{N,f,\hat{x}_0}(\hat{x}).\end{aligned}$$

Es decir, que $P_N(\hat{x}) = P_{N,f,\hat{x}_0}(\hat{x})$ para toda $\hat{x} \in \mathbb{R}^n$ que es lo que queríamos demostrar. ■

Los polinomios de Taylor son una herramienta muy importante en el análisis del comportamiento de una función alrededor de un punto, y nos serán muy útiles cuando en la próxima sección abordemos el tema de los valores máximos y mínimos de una función. Por ahora, veremos con un ejemplo cómo se pueden usar para el cálculo de límites.

Ejemplo 4.14 *Considere la función*

$$f(x, y) = \frac{\ln(1 + x^2) + \ln(1 + y^2)}{x^2 + y^2}$$

para $(x,y) \in \mathbb{R}^2$, $(x,y) \neq (0,0)$.

Nuestro objetivo es calcular

$$\lim_{(x,y)\to(0,0)} f(x,y).$$

Para ello, bastará con calcular el polinomio de Taylor de grado 2 *en el* $(0,0)$ *de la función*

$$g(x,y) = \ln(1+x^2) + \ln(1+y^2),$$

dado que el denominador de f *es* $x^2+y^2 = \|(x,y)\|^2$.

Para la función g *se tiene que*

$$\frac{\partial g}{\partial x}(x,y) = \frac{2x}{1+x^2} \qquad y \qquad \frac{\partial g}{\partial y}(x,y) = \frac{2y}{1+y^2},$$

de modo que

$$\frac{\partial^2 g}{\partial x^2}(x,y) = \frac{2}{1+x^2} - \frac{4x^2}{(1+x^2)^2} \qquad y \qquad \frac{\partial^2 g}{\partial y \partial x}(x,y) = 0.$$

Análogamente,

$$\frac{\partial^2 g}{\partial x \partial y}(x,y) = 0 \qquad y \qquad \frac{\partial^2 g}{\partial y^2}(x,y) = \frac{2}{1+y^2} - \frac{4y^2}{(1+y^2)^2}.$$

De esta forma, tenemos que

$$\begin{aligned} P_{2,g,\hat{0}}(x,y) &= g(\hat{0}) + \frac{\partial g}{\partial x}(\hat{0})x + \frac{\partial g}{\partial y}(\hat{0})y \\ &\quad + \frac{1}{2}\left(\frac{\partial^2 g}{\partial x^2}(\hat{0})x^2 + \frac{\partial^2 g}{\partial y \partial x}(\hat{0})xy + \frac{\partial^2 g}{\partial x \partial y}(\hat{0})yx + \frac{\partial^2 g}{\partial y^2}(\hat{0})y^2\right) \\ &= x^2 + y^2 \end{aligned}$$

y, por lo tanto, que

$$\begin{aligned}
\lim_{(x,y)\to(0,0)} f(x,y) &= \lim_{(x,y)\to(0,0)} \frac{\ln(1+x^2)+\ln(1+y^2)}{x^2+y^2} \\
&= \lim_{(x,y)\to(0,0)} \frac{g(x,y)}{x^2+y^2} \\
&= \lim_{(x,y)\to(0,0)} \frac{P_{2,g,\hat{0}}(x,y)+R_{2,g,\hat{0}}(x,y)}{x^2+y^2} \\
&= \lim_{(x,y)\to(0,0)} \frac{x^2+y^2+R_{2,g,\hat{0}}(x,y)}{x^2+y^2} \\
&= \lim_{(x,y)\to(0,0)} \left(1+\frac{R_{2,g,\hat{0}}(x,y)}{\|(x,y)\|^2}\right) \\
&= 1+0 \\
&= 1.
\end{aligned}$$

4.6. Máximos y mínimos

Una de las aplicaciones más comunes de la derivada es la localización de los puntos en los que una función alcanza sus valores máximo y mínimo. De acuerdo con el corolario 2.3 del capítulo 2, si una función $f : A \subset \mathbb{R}^n \to \mathbb{R}$ es continua en un subconjunto B de su dominio A, y este subconjunto es cerrado y acotado (es decir compacto), sólo en este caso podemos estar seguros de que existen un par de puntos en B en los cuales la función f alcanza su valor máximo y su valor mínimo, en donde además estos valores son máximo y mínimo sólo con respecto a los valores de f sobre los elementos del subconjunto B.

Puesto que todo conjunto $B \subset \mathbb{R}^n$ que sea cerrado se puede "descomponer" en la unión de su interior y su frontera ($B = \mathrm{Fr}(B) \cup \mathrm{int}(B)$), que son dos conjuntos ajenos, y dado que $\mathrm{int}(B)$ siempre es un conjunto abierto y $\mathrm{Fr}(B)$ siempre es un conjunto cerrado (y "flaco", ya que por el problema 15 del capítulo 1 se tiene que $\mathrm{int}(\mathrm{Fr}(B)) = \varnothing$), vamos a "reducir" nuestro análisis a estos dos casos: cuando los puntos en los que una función f alcanza

su valor máximo o mínimo (o alguno de ellos) pertenecen al interior de B, y cuando estos mismos puntos (o alguno de ellos) pertenecen a la frontera de B.

De acuerdo con el enfoque anterior, lo que haremos será analizar en general qué características específicas satisfacen los puntos en los que una función f alcanza su valor máximo y su valor mínimo, dependiendo de si éstos pertenecen a un conjunto abierto (como el $\text{int}(B)$) o a un conjunto cerrado con interior vacío (como la $\text{Fr}(B)$, cuando B es cerrado).

Definición 4.12 *Sean $f : A \subset \mathbb{R}^n \to \mathbb{R}$, $B \subset A$ y $\hat{x}_0 \in B$. Decimos que:*

1. *f alcanza (o tiene) un valor máximo (mínimo) en $\hat{x}_0$ sobre B si $f(\hat{x}) \leq f(\hat{x}_0)$ $(f(\hat{x}_0) \leq f(\hat{x}))$ para todo $\hat{x} \in B$.*

2. *f alcanza (o tiene) un valor máximo local (mínimo local) en $\hat{x}_0$ sobre B si existe $r > 0$ tal que $f(\hat{x}) \leq f(\hat{x}_0)$ $(f(\hat{x}_0) \leq f(\hat{x}))$ para todo $\hat{x} \in B \cap B_r(\hat{x}_0)$.*

Con respecto a las definiciones anteriores es importante hacer dos observaciones. Una es que, necesariamente, si en un punto $\hat{x}_0$ una función f alcanza un valor máximo (mínimo) sobre un conjunto B, entonces también alcanza (o tiene) un valor máximo local (mínimo local) sobre B, pero no recíprocamente. Es decir, si f alcanza en $\hat{x}_0$ un valor máximo local (mínimo local) sobre B, este no es necesariamente máximo (mínimo) sobre todo B (ver figura 4.12). La segunda observación es que, si B es un conjunto abierto, entonces f alcanza en $\hat{x}_0$ un valor máximo local (mínimo local) sobre B si y sólo si existe $r > 0$ tal que $f(\hat{x}) \leq f(\hat{x}_0)$ $(f(\hat{x}_0) \leq f(\hat{x}))$ para todo $\hat{x} \in B_r(\hat{x}_0) \subset B$.

El primer resultado importante que vamos a probar en esta sección tiene que ver con los puntos en los que una función f alcanza un valor máximo (o mínimo) local sobre un conjunto abierto. Si en un punto $\hat{x}_0$, además de satisfacerse la propiedad anterior, f también es derivable, entonces la derivada en $\hat{x}_0$ $(Df(\hat{x}_0))$ tiene que ser la función lineal constante 0 (como seguramente el lector ya está observando, este resultado generaliza lo que sucede para las funciones de $\mathbb{R}$ en $\mathbb{R}$).

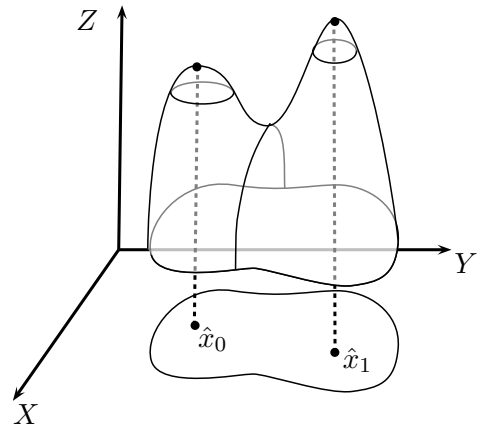


Figura 4.12: En el punto $\hat{x}_0$ la función f tiene un máximo local pero no global, y en el punto $\hat{x}_1$ tiene un máximo local que también es global.

Proposición 4.19 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$ (U abierto) y $\hat{x}_0 \in U$. Si f alcanza en $\hat{x}_0$ un valor máximo local (mínimo local) sobre U y f es derivable en $\hat{x}_0$, entonces $Df(\hat{x}_0)$ es la constante cero ($Df(\hat{x}_0) \equiv 0$).*

Demostración. Dado que la derivada de f en $\hat{x}_0$ está representada (en un sistema coordenado ortonormal) por la matriz de $1 \times n$ cuya $i-$ésima entrada está dada por $\frac{\partial f}{\partial x_i}(\hat{x}_0)$, bastará con demostrar que cada uno de éstos números es 0.

Suponiendo que f alcanza en $\hat{x}_0$ un valor máximo local sobre U (el caso del mínimo local se prueba de manera análoga), sea $r > 0$ tal que $f(\hat{x}) \leq f(\hat{x}_0)$ para todo $\hat{x} \in B_r(\hat{x}_0) \subset U$. Por tanto, si $|h| < r$ tenemos que $f(\hat{x}_0 + h\hat{e}_i) \leq f(\hat{x}_0)$, de tal forma que

$$f(\hat{x}_0 + h\hat{e}_i) - f(\hat{x}_0) \leq 0.$$

Ahora, si tomamos h tal que $0 < h < r$, entonces

$$\frac{f(\hat{x}_0 + h\hat{e}_i) - f(\hat{x}_0)}{h} \leq 0$$

y por lo tanto

$$\begin{aligned}\frac{\partial f}{\partial x_i}(\hat{x}_0) &= \lim_{h\to 0}\frac{f(\hat{x}_0+h\hat{e}_i)-f(\hat{x}_0)}{h}\\ &\leq 0.\end{aligned}$$

Por otra parte, si tomamos h tal que $-r<h<0$, entonces

$$0\leq \frac{f(\hat{x}_0+h\hat{e}_i)-f(\hat{x}_0)}{h},$$

de donde

$$\begin{aligned}0 &\leq \lim_{h\to 0}\frac{f(\hat{x}_0+h\hat{e}_i)-f(\hat{x}_0)}{h}\\ &= \frac{\partial f}{\partial x_i}(\hat{x}_0).\end{aligned}$$

Entonces se tiene que

$$0\leq \frac{\partial f}{\partial x_i}(\hat{x}_0)\leq 0$$

y por lo tanto

$$\frac{\partial f}{\partial x_i}(\hat{x}_0)=0.$$

■

Como seguramente el lector ya habrá notado, si bien la proposición anterior es importante, ésta sólo nos proporciona una consecuencia (o condición) necesaria del hecho de que una función alcance en $\hat{x}_0$ un valor máximo (o mínimo) local. Como suele suceder con frecuencia, esta condición no es suficiente. En efecto, el hecho de que la derivada de una función f en un punto $\hat{x}_0$ sea la constante 0, no es suficiente para que podamos asegurar que f alcance en $\hat{x}_0$ un valor máximo (o mínimo) local.

Antes de dar un ejemplo que ilustra este hecho, mencionaremos que a pesar de todo, aquellos puntos en que la derivada de una función es la constante 0 resultan ser muy importantes y reciben un nombre especial: *puntos críticos*, lo cual dejamos plasmado en la siguiente

Definición 4.13 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $\hat{x}_0 \in U$. Si f es derivable en $\hat{x}_0$ y la derivada de f en $\hat{x}_0$ ($Df(\hat{x}_0)$) es la constante 0, decimos que $\hat{x}_0$ es un punto crítico de f.*

Con base en la definición anterior, lo que mostraremos en el siguiente ejemplo es que los puntos críticos de una función f no tienen por qué ser necesariamente puntos en los que f alcance un valor máximo (o mínimo) local.

Ejemplo 4.15 *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ dada por la expresión*

$$f(x, y) = x^2 - y^2.$$

Dado que

$$\frac{\partial f}{\partial x}(x, y) = 2x \qquad y \qquad \frac{\partial f}{\partial y}(x, y) = -2y,$$

se tiene que el único punto crítico de f es el $(0, 0)$.

Sin embargo, para cualquier $r > 0$ se tiene que: si $0 < |x| < r$, entonces $(x, 0) \in B_r((0, 0))$ y

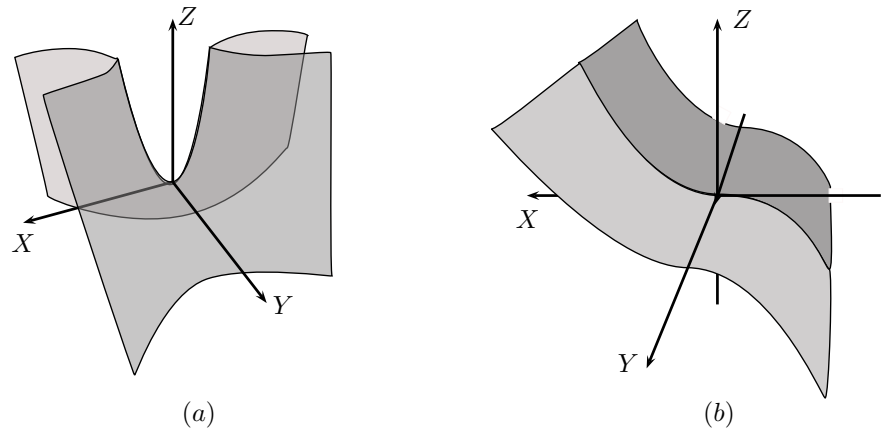
$$\begin{aligned} f(x, 0) &= x^2 \\ &> 0 \\ &= f(0, 0). \end{aligned}$$

Por otra parte, si $0 < |y| < r$, entonces $(0, y) \in B_r((0, 0))$ y

$$\begin{aligned} f(0, y) &= -y^2 \\ &< 0 \\ &= f(0, 0), \end{aligned}$$

lo que demuestra que f no tiene un máximo local ni un mínimo local en el $(0, 0)$ (ver figura 4.13 (a)).

Aquellos puntos críticos en los que una función no tiene un máximo local o un mínimo local, justo por el aspecto geométrico del ejemplo anterior en una vecindad del $(0,0)$ (el de una silla de montar), reciben el nombre de *punto silla*, aunque no todos los puntos silla tienen este aspecto. Por ejemplo para la función $f(x,y) = x^3$, todos sus puntos críticos (los puntos de la forma $(0,y)$) son puntos silla, en el sentido de que en ellos la función f no tiene un máximo local o un mínimo local. Sin embargo, geométricamente la función f no tiene el aspecto de una silla de montar (ver figura 4.13 (b)).



(a) (b)

Figura 4.13: La función $f(x,y) = x^2 - y^2$ (figura (a)) tiene un punto silla en el punto $(0,0)$, y la función $f(x,y) = x^3$ (figura (b)) tiene puntos silla en los puntos de la forma $(0,y)$.

El ejemplo anterior, y el hecho de que la proposición 4.19 es válida para máximos o mínimos locales nos plantea el problema de contar con un criterio que nos permita "clasificar" a los puntos críticos de una función.

Lo que haremos a continuación será desarrollar un criterio análogo al criterio de la segunda derivada para la clasificación de puntos críticos de funciones de $\mathbb{R}$ en $\mathbb{R}$, pero para ello será necesario introducir el concepto (que tomaremos "prestado" nuevamente del ¡Álgebra Lineal!) de: *forma cuadrática*. Más adelante haremos una breve revisión de este concepto.

De acuerdo con el problema 36 (que es una consecuencia del teorema de Taylor), si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es una función de clase C^2 en U, $\hat{x}_0 \in U$ y $r > 0$ son tales que $B_r(\hat{x}_0) \subset U$, sabemos que si $\hat{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$ es tal que $\left\|\hat{h}\right\| < r$, entonces

$$\begin{aligned} f(\hat{x}_0 + \hat{h}) &= P_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}) + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}) \\ &= f(\hat{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0)h_i + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0)h_i h_j + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}), \end{aligned}$$

en donde

$$\lim_{\hat{h}\to\hat{0}} \frac{R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2} = 0.$$

De este modo, si $\hat{x}_0$ es un punto crítico de f, se tiene que

$$f(\hat{x}_0 + \hat{h}) - f(\hat{x}_0) = \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0)h_i h_j + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}).$$

De esta última identidad, nos detendremos a estudiar con detalle a la expresión

$$\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0)h_i h_j \tag{4.42}$$

y lo primero que haremos notar es que se puede escribir en forma matricial, de la siguiente manera

$$\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0)h_i h_j = \begin{bmatrix} h_1 & \cdots & h_n \end{bmatrix} A \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix},$$

en donde la matriz A es

$$A = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\hat{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\hat{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\hat{x}_0) \end{bmatrix}. \tag{4.43}$$

La ventaja de escribir a la expresión 4.42 en forma matricial es que nos permite relacionarla con un concepto que ya mencionamos: las formas cuadráticas.

Las formas cuadráticas en $\mathbb{R}^n$ son (además de la función constante 0, que es la única forma cuadrática que es constante), las funciones polinomiales de las coordenadas $x_1, \ldots, x_n$ que tienen la particularidad de que todos los monomios que la forman son expresiones de grado 2. Por ejemplo, en el caso de $\mathbb{R}^2$, cualquier forma cuadrática en las variables x, y se escribe como

$$ax^2 + 2bxy + cy^2 \tag{4.44}$$

la cual, como en el caso de la expresión 4.42, se puede escribir en forma matricial de la siguiente manera:

$$\begin{aligned} ax^2 + 2bxy + cy^2 &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x & y \end{bmatrix}^t \\ &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \end{aligned}$$

En general, toda forma cuadrática en $\mathbb{R}^n$ se identifica con una matriz A de $n \times n$ con entradas reales ($A \in M_{2\times 2}(\mathbb{R})$) la cual es simétrica, es decir, que es igual a su transpuesta ($A = A^t$). Y recíprocamente, si A es una de estas matrices, la función $F : \mathbb{R}^n \to \mathbb{R}$ definida como

$$\begin{aligned} F(x_1, \ldots, x_n) &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} A \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^t \\ &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \end{aligned}$$

claramente define una forma cuadrática en $\mathbb{R}^n$.

A la matriz A dada por 4.43 se le conoce como la *matriz hessiana*[3] de f en $\hat{x}_0$, que de aquí en adelante denotaremos por $Hf(\hat{x}_0)$, y a la forma

[3]Llamada así en honor de Ludwig Otto Hesse (22 abril 1811 - 4 agosto 1874) quien fue un matemático alemán. Hesse nació en K$'$onigsberg, Prussia, y murió en Munich, Bavaria. Trabajó en la teoría de invariantes.

cuadrática asociada a ésta le llamaremos *la hessiana de f en $\hat{x}_0$*. La hessiana de f en $\hat{x}_0$ la denotaremos por $Hf_{\hat{x}_0}$ de tal forma que si $\hat{x} = (x_1, \ldots, x_n)$, entonces

$$\begin{aligned} Hf_{\hat{x}_0}(\hat{x}) &= Hf_{\hat{x}_0}(x_1, \ldots, x_n) \\ &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} Hf(\hat{x}_0) \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^t \\ &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} Hf(\hat{x}_0) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \end{aligned}$$

Hay una clasificación de las formas cuadráticas que en particular resultará muy importante para el problema de la identificación de los puntos críticos de una función. Se dice que la forma cuadrática $F : \mathbb{R}^n \to \mathbb{R}$ es *semipositiva* (análogamente *seminegativa*) si $F(\hat{x}) \geq 0$ ($F(\hat{x}) \leq 0$) para toda $\hat{x} \in \mathbb{R}^n$ y si F satisface la condición de que $F(\hat{x}) = 0$ si y sólo si $\hat{x} = \hat{0}$, entonces decimos que F es *no degenerada*.

Lo interesante de esta clasificación de las formas cuadráticas es que, con base en ella, podremos establecer criterios que nos permitan determinar si los puntos críticos de una función son puntos en los que ésta alcanza un valor máximo o mínimo, locales en ambos casos.

En efecto, lo siguiente que haremos será probar un resultado en el que, dependiendo de qué tipo de forma cuadrática resulte ser la hessiana de f en un punto $\hat{x}_0$, podremos asegurar que f tiene un cierto tipo de valor extremo (máximo o mínimo local). Para ello, será necesario probar antes un sencillo resultado que formularemos en el siguiente

Lema 4.1 *Si $F : \mathbb{R}^n \to \mathbb{R}$ es una forma cuadrática semipositiva (seminegativa) y no degenerada asociada a una matriz A (es decir que $F(\hat{x}) = \hat{x}A\hat{x}^t$), entonces existe $M > 0$ ($m < 0$) tal que*

$$F(\hat{h}) \geq M \left\|\hat{h}\right\|^2 \qquad \left(F(\hat{h}) \leq m \left\|\hat{h}\right\|^2\right)$$

para toda $\hat{h} \in \mathbb{R}^n$.

Demostración. Dado que $S^{n-1} = \{\hat{x} \in \mathbb{R}^n \mid \|\hat{x}\| = 1\}$ es un conjunto cerrado y acotado (compacto) y toda forma cuadrática es una función continua, por el corolario 2.3 del capítulo 2 sabemos que existe un valor mínino para F (si F es semipositiva), que llamaremos $M > 0$ (un valor máximo para F, si F es seminegativa, que llamaremos $m < 0$), tal que

$$F(\hat{x}) \geq M \qquad (F(\hat{x}) \leq m)$$

para toda $\hat{x} \in \mathbb{R}^n$.

Por tanto, dado que las desigualdades que deseamos probar se satisfacen si $\hat{h} = \hat{0}$, si tomamos $\hat{h} \neq \hat{0}$ y hacemos $\hat{x} = \hat{h}/\left\|\hat{h}\right\|$, entonces se tiene que $\hat{x} \in S^{n-1}$ y por lo tanto

$$\begin{aligned} M &\leq F(\hat{x}) \\ &= \hat{x}A\hat{x}^t \\ &= \left(\frac{\hat{h}}{\left\|\hat{h}\right\|}\right) A \left(\frac{\hat{h}}{\left\|\hat{h}\right\|}\right)^t \\ &= \frac{1}{\left\|\hat{h}\right\|^2}\left(\hat{h}A\hat{h}^t\right) \\ &= \frac{1}{\left\|\hat{h}\right\|^2}F(\hat{h}). \end{aligned}$$

Es decir, se tiene que

$$F(\hat{h}) \geq M\left\|\hat{h}\right\|^2$$

para toda $\hat{h} \in \mathbb{R}^n$ (análogamente se prueba que $F(\hat{h}) \leq m\left\|\hat{h}\right\|^2$ para toda $\hat{h} \in \mathbb{R}^n$). ■

Una vez hecho todo lo anterior, estamos en condiciones de formular un resultado que nos permitirá determinar si un punto crítico $\hat{x}_0$ de una función f es un punto en donde ésta alcanza un máximo local o un mínimo local.

Sólo probaremos uno de estos casos y como es de imaginar, el otro caso quedará como un problema para el lector.

Proposición 4.20 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^2 en U, y $\hat{x}_0 \in U$ un punto crítico de f. Se satisface lo siguiente:*

1. *si la hessiana de f en $\hat{x}_0$ es una forma cuadrática semipositiva y no degenerada, entonces f tiene un mínimo local en $\hat{x}_0$, y*

2. *si la hessiana de f en $\hat{x}_0$ es una forma cuadrática seminegativa y no degenerada, entonces f tiene un máximo local en $\hat{x}_0$.*

Demostración. Sólo probaremos el inciso 1. Sea $r_0 > 0$ tal que $B_{r_0}(\hat{x}_0) \subset U$. Como ya habíamos mencionado anteriormente, por el problema 36 sabemos que, como $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es una función de clase C^2 en U y $\hat{x}_0 \in U$ es un punto crítico de f, si $\hat{h} = (h_1, \ldots, h_n) \in \mathbb{R}^2$ es tal que $\left\|\hat{h}\right\| < r_0$, entonces

$$\begin{aligned} & f(\hat{x}_0 + \hat{h}) \\ & \quad = P_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}) + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}) \\ & \quad = f(\hat{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0) h_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0) h_i h_j + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}) \\ & \quad = f(\hat{x}_0) + \frac{1}{2} H f_{\hat{x}_0}(\hat{h}) + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h}), \end{aligned}$$

en donde

$$\lim_{\hat{h} \to \hat{0}} \frac{R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2} = 0. \tag{4.45}$$

Ahora, dado que $Hf_{\hat{x}_0}$ es semipositiva, por el lema 4.1 sabemos que existe $M > 0$ tal que

$$Hf_{\hat{x}_0}(\hat{h}) \geq M \left\|\hat{h}\right\|^2$$

para toda $\hat{h} \in \mathbb{R}^n$. Por tanto, por 4.45 sabemos que existe $\delta > 0$ tal que si $\left\|\hat{h}\right\| < \delta$, entonces

$$\frac{\left|R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})\right|}{\left\|\hat{h}\right\|^2} < \frac{M}{2}$$

y en particular, que

$$0 < \frac{M}{2} + \frac{R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2}.$$

Por lo tanto, si $r = \text{mín}\{\delta, r_0\}$ y $0 < \left\|\hat{h}\right\| < r$, tendremos que

$$\begin{aligned}
\frac{f(\hat{x}_0 + \hat{h}) - f(\hat{x}_0)}{\left\|\hat{h}\right\|^2} &= \frac{\frac{1}{2}Hf_{\hat{x}_0}(\hat{h}) + R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2} \\
&= \frac{1}{2}\frac{Hf_{\hat{x}_0}(\hat{h})}{\left\|\hat{h}\right\|^2} + \frac{R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2} \\
&\geq \frac{M}{2} + \frac{R_{2,f,\hat{x}_0}(\hat{x}_0 + \hat{h})}{\left\|\hat{h}\right\|^2} \\
&> 0,
\end{aligned}$$

de donde concluimos que

$$f(\hat{x}_0) \leq f(\hat{x}_0 + \hat{h})$$

si $\left\|\hat{h}\right\| < r$, lo que demuestra que f tiene un mínimo local en $\hat{x}_0$. ■

Como el lector habrá notado, el resultado anterior es una especie de "equivalente" (en $\mathbb{R}^n$) al "criterio de la segunda derivada" para la clasificación de puntos críticos de funciones de $\mathbb{R}$ en $\mathbb{R}$, y como sucede en ese caso, su "recíproco" no se cumple, lo que ilustramos en el siguiente

Ejemplo 4.16 *Sea* $f : \mathbb{R}^2 \to \mathbb{R}$ *definida como*

$$f(x, y) = x^4 + y^4$$

para cada $(x, y) \in \mathbb{R}^2$*. Como es fácil de verificar (ver figura 4.14), el* $(0, 0)$ *es el punto en el que* f *alcanza su mínimo global (y por tanto también local) y dado que*

$$\frac{\partial^2 f}{\partial x^2}(0, 0) = \frac{\partial^2 f}{\partial y^2}(0, 0) = \frac{\partial^2 f}{\partial y \partial x}(0, 0) = 0$$

se tiene que la hessiana de f *en el* $(0, 0)$ *es la función constante* 0*, la cual es la más degenerada (¡en el sentido matemático de la palabra!) de todas las formas cuadráticas.*

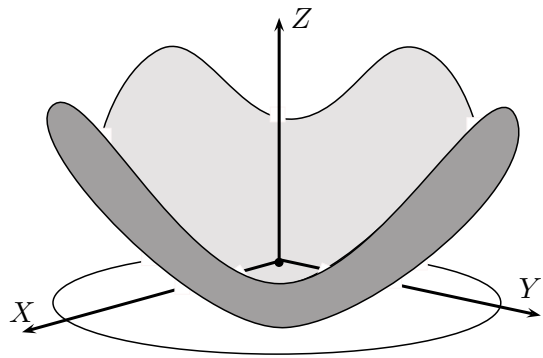


Figura 4.14: La función $f(x, y) = x^4 + y^4$ tiene un mínimo global (y por tanto también local) en el punto $(0, 0)$, y su forma cuadrática asociada en ese punto (su hessiana) es degenerada.

A pesar del ejemplo anterior, para la proposición 4.20 se puede establecer una proposición "casi-recíproca". Como seguramente el lector recordará, para el caso de funciones de $\mathbb{R}$ en $\mathbb{R}$ se sabe que si f tiene un máximo (mínimo) local en un punto x_0, y $f''(x_0)$ existe, entonces se debe cumplir que $f''(x_0) \leq 0$ $(f''(x_0) \geq 0)$ (resultado que por cierto nos será de mucha utilidad en la prueba de nuestra siguiente proposición).

Para una función f de $\mathbb{R}^n$ en $\mathbb{R}$ se "extiende" el resultado anterior en el sentido de que, si f tiene un máximo (o un mínimo) local en un punto $\hat{x}_0$, entonces la hessiana de f en $\hat{x}_0$ será seminegativa (o semipositiva). Es decir, lo único que no podremos asegurar es que la hessiana es no degenerada (así como en el caso real no podemos asugurar que $f''(x_0)$ no sea 0).

Proposición 4.21 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^2 en U, y $\hat{x}_0 \in U$ un punto crítico de f. Se satisface lo siguiente:*

1. *si f tiene un mínimo local en $\hat{x}_0$, entonces la hessiana de f en $\hat{x}_0$ es una forma cuadrática semipositiva, y*

2. *si f tiene un máximo local en $\hat{x}_0$, entonces la hessiana de f en $\hat{x}_0$ es una forma cuadrática seminegativa.*

Demostración. Sólo probaremos el inciso 2. Sean, $r_0 > 0$ tal que $B_{r_0}(\hat{x}_0) \subset U$, y $\hat{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Definimos $\gamma : (-r_0, r_0) \subset \mathbb{R} \to \mathbb{R}^n$ como $\gamma(t) = \hat{x}_0 + t\hat{u}$ y $g : (-r_0, r_0) \subset \mathbb{R} \to \mathbb{R}$ como $g = f \circ \gamma$.

Nótese que, por la proposición 4.16 sabemos que g es derivable en su dominio y que

$$\begin{aligned} g'(t) &= \nabla f(\gamma(t)) \cdot \gamma'(t) \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\gamma(t)) u_i. \end{aligned}$$

Dado que f es de clase C^2 en U, por la misma proposición 4.16 (aplicada a cada miembro de la suma anterior), se tiene

$$\begin{aligned} g''(t) &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\gamma(t)) u_j \right) u_i \\ &= \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\gamma(t)) u_i u_j, \end{aligned}$$

de tal forma que

$$\begin{aligned} g''(0) &= \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\gamma(0))u_i u_j \\ &= \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0)u_i u_j \\ &= Hf_{\hat{x}_0}(\hat{u}). \end{aligned}$$

Por otra parte, es sencillo demostrar (problema 43) que, si f tiene un máximo local en $\hat{x}_0$, entonces g tiene un máximo local en $t = 0$, de tal forma que, por el resultado para funciones de $\mathbb{R}$ en $\mathbb{R}$ que mencionamos previamente a esta proposición, se debe tener que

$$\begin{aligned} Hf_{\hat{x}_0}(\hat{u}) &= g''(0) \\ &\leq 0. \end{aligned}$$

Por lo tanto, si $\hat{x} \in \mathbb{R}^n$, $\hat{x} \neq \hat{0}$, y hacemos $\hat{u} = \hat{x}/\left\|\hat{x}\right\|$, se tendrá que

$$\begin{aligned} Hf_{\hat{x}_0}(\hat{u}) &= Hf_{\hat{x}_0}\left(\frac{\hat{x}}{\|\hat{x}\|}\right) \\ &= \frac{1}{\|\hat{x}\|^2} Hf_{\hat{x}_0}(\hat{x}) \\ &\leq 0, \end{aligned}$$

de donde $Hf_{\hat{x}_0}(\hat{x}) \leq 0$ para toda $\hat{x} \in \mathbb{R}^n$, lo que prueba que la hessiana de f en $\hat{x}_0$ es seminegativa. ■

Aun cuando todo lo desarrollado hasta aquí parece funcionar muy bien, no podemos dejar de mencionar que el problema de determinar si una función f tiene un máximo o un mínimo local (o ninguno de estos dos) en un punto crítico $\hat{x}_0$, se "trasladadó" al problema de determinar si la hessiana de f en este punto $\hat{x}_0$ es una función cuadrática semipositiva o seminegativa (o ninguna de estas dos).

En la siguiente sección mencionaremos (sin probar) algunas condiciones necesarias y suficientes para que una forma cuadrática F en $\mathbb{R}^n$ (en términos de su matriz asociada) sea semipositiva o seminegativa, y no degenerada. Como ya mencionamos, esstas condiciones necesarias y suficientes son un tema importante de Álgebra Lineal que desafortunadamente no podemos incluir en este texto (pero que el lector interesado puede consultar en [2]) para saber más de él).

Sin embargo, para que el amable lector no se sienta muy desalentado por esta ausencia, mostraremos y probaremos cuáles son estas condiciones para el caso particular de las formas cuadráticas en $\mathbb{R}^2$, las que dejaremos plasmadas en la siguiente

Proposición 4.22 *Sea $F : \mathbb{R}^2 \to \mathbb{R}$ la forma cuadrática asociada a la matriz*

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

y que está dada por la expresión

$$\begin{aligned} F(x,y) &= \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x & y \end{bmatrix}^t \\ &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= ax^2 + 2bxy + cy^2. \end{aligned}$$

F es semipositiva (seminegativa) y no degenerada si y sólo si $a > 0$ ($a < 0$) y $ac - b^2 = \det(A) > 0$.

Demostración. Supongamos que F es semipositiva (el caso en que sea seminegativa se prueba de manera análoga) y no degenerada. Por esta última propiedad, dado que

$$a = F(1,0) > 0,$$

concluimos que $a > 0$. Por tanto, podemos reescribir a F como

$$\begin{aligned} F(x,y) &= ax^2 + 2bxy + cy^2 \\ &= a\left(x + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2 \\ &= a\left(x + \frac{b}{a}y\right)^2 + \left(\frac{ac - b^2}{a}\right)y^2, \end{aligned} \tag{4.46}$$

de tal forma que

$$\begin{aligned} 0 &< F(b/a, -1) \\ &= \frac{ac - b^2}{a} \end{aligned}$$

y por lo tanto $ac - b^2 > 0$.

Supongamos ahora que $a > 0$ y $ac - b^2 > 0$. De la identidad 4.46 concluimos que $F(x,y) \geq 0$ para toda $(x,y) \in \mathbb{R}^2$, es decir, F es semipositiva.

Finalmente, si (x,y) es tal que $F(x,y) = 0$, se debe tener que

$$\left(\frac{ac - b^2}{a}\right)y^2 = 0$$

y

$$\left(x + \frac{b}{a}y\right)^2 = 0,$$

de tal forma que de la primera identidad concluimos que $y = 0$, y sustituyendo esto último en la segunda identidad, concluimos que $x = 0$. Esto prueba que F es no degenerada. ■

Como una consecuencia inmediata de la proposición anterior y de la proposición 4.20, podemos formular el siguiente resultado, que establece criterios muy concretos para determinar si un punto crítico de una función de $\mathbb{R}^2$ en $\mathbb{R}$ es un máximo o mínimo local.

Proposición 4.23 *Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ de clase C^2 en U, y $\hat{x}_0 \in U$ un punto crítico de f.*

1. *Si*

$$\frac{\partial^2 f}{\partial x^2}(\hat{x}_0) > 0 \qquad y \qquad \frac{\partial^2 f}{\partial x^2}(\hat{x}_0)\frac{\partial^2 f}{\partial y^2}(\hat{x}_0) - \left(\frac{\partial^2 f}{\partial y \partial x}(\hat{x}_0)\right)^2 > 0,$$

entonces f tiene un mínimo local en $\hat{x}_0$, y

2. *si*

$$\frac{\partial^2 f}{\partial x^2}(\hat{x}_0) < 0 \qquad y \qquad \frac{\partial^2 f}{\partial x^2}(\hat{x}_0)\frac{\partial^2 f}{\partial y^2}(\hat{x}_0) - \left(\frac{\partial^2 f}{\partial y \partial x}(\hat{x}_0)\right)^2 > 0,$$

entonces f tiene un máximo local en $\hat{x}_0$.

Antes de hacer unos breves comentarios adicionales acerca de las formas cuadráticas, daremos algunos ejemplos que ilustren el trabajo realizado hasta ahora sobre el tema de máximos y mínimos.

Ejemplo 4.17

1. *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ dada por*

$$f(x, y) = (x + 1)^2 + y^2$$

y $B = \{(x, y) \in \mathbb{R}^2 \mid \|(x, y)\| \leq 2\}$.

De acuerdo con lo realizado hasta este momento, si los puntos en los que f alcanza un valor máximo o mínimo local (o global) sobre B, se encuentran en el interior de B, entonces deben ser puntos críticos de f, de modo que lo primero que se tendrá que hacer será localizar este tipo de puntos.

Para lograr esto, hay que resolver las ecuaciones

$$\frac{\partial f}{\partial x}(x, y) = 0 \qquad y \qquad \frac{\partial f}{\partial y}(x, y) = 0,$$

que en este caso son

$$2(x + 1) = 0 \qquad y \qquad 2y = 0,$$

de donde obtenemos que $(-1,0)$ *es el único punto crítico de* f *(en todo* $\mathbb{R}^2$*).*

Por otra parte, dado que $f(x,y) \geq 0$ *para toda* $(x,y) \in \mathbb{R}^2$*, que* $f(-1,0) = 0$ *y* $(-1,0) \in B$*, sin necesidad de mayores cálculos, podemos concluir que* f *alcanza su valor mínimo sobre* B *(de hecho, sobre todo* $\mathbb{R}^2$*) en el punto* $(-1,0)$ *y que dicho valor mínimo es* 0*.*

En cuanto al valor máximo de f *sobre* B*, dado que* f *no tiene más puntos críticos, este valor máximo se debe de alcanzar en algún punto de la frontera de* B*, es decir, sobre el conjunto*

$$\mathrm{Fr}(B) = \left\{(x,y) \in \mathbb{R}^2 \mid \|(x,y)\| = 2\right\}.$$

Dado que para este tipo de conjuntos (cerrados con interior vacío) no hemos desarrollado ninguna herramienta que nos permita resolver este problema, para este caso particular recurriremos al siguiente hecho: la $\mathrm{Fr}(B)$ *se puede obtener como la imagen de una función de* $\mathbb{R}$ *en* $\mathbb{R}^2$*. En efecto, nótese que la función* $\gamma : [0,2\pi] \subset \mathbb{R} \to \mathbb{R}^2$ *dada por*

$$\gamma(t) = (2\cos(t), 2\,\mathrm{sen}(t))$$

es una parametrización del conjunto $\mathrm{Fr}(B)$*, es decir,* $\gamma([0,2\pi]) = \mathrm{Fr}(B)$*.*

Aprovechando este hecho, si definimos la función $g : [0,2\pi] \subset \mathbb{R} \to \mathbb{R}$ *como*

$$\begin{aligned} g(t) &= (f \circ \gamma)(t) \\ &= f(\gamma(t)) \\ &= (2\cos(t) + 1)^2 + 4\,\mathrm{sen}^2(t) \\ &= 4\cos(t) + 5 \end{aligned}$$

y encontramos los puntos del dominio de g *en los que esta función alcanza sus valores máximo y mínimo, podremos entonces localizar los valores máximo y mínimo de* f *sobre la* $\mathrm{Fr}(B)$*.*

Como ya se sabe, hay que proceder de manera análoga a lo que estamos haciendo con f; es decir, hay que localizar los puntos críticos de g en el intervalo abierto $(0, 2\pi)$, determinar si en estos puntos críticos g alcanza un máximo o mínimo local, y comparar el valor de g en estos puntos con los valores de g en los extremos del intervalo $[0, 2\pi]$ (que son la frontera de $[0, 2\pi]$).

De esta forma, lo primero que hay que localizar son los valores de $t \in (0, 2\pi)$ para los cuales se cumple que

$$\begin{aligned} g'(t) &= -4\,\mathrm{sen}(t) \\ &= 0 \end{aligned}$$

y que claramente sólo sucede para $t = \pi$. Dado que $g(t) \geq 1$ para toda $t \in [0, 2\pi]$ y que $g(\pi) = 1$, concluimos que g alcanza su valor mínimo en $t = \pi$ y que este valor mínimo es 1.

Finalmente, dado que $g(0) = g(2\pi) = 9$, concluimos que el valor máximo de g sobre el intervalo $[0, 2\pi]$ es 9, y que este valor lo alcanza en $t = 0$ y $t = 2\pi$ (es decir, en la frontera de su dominio). De todo lo anterior concluimos las siguiente dos cosas: una, que el valor mínimo de f sobre $\mathrm{Fr}(B)$ es 1 y que este valor lo alcanza en el punto $\gamma(\pi) = (-2, 0)$; dos, que su valor máximo sobre la $\mathrm{Fr}(B)$ lo alcanza en el punto $\gamma(0) = \gamma(2\pi) = (2, 0)$ y que este valor máximo es 9.

Por lo tanto, comparando los valores de f en los puntos críticos que están en el interior de B ($f(-1, 0) = 0$), junto con los valores máximo y mínimo de f sobre la $\mathrm{Fr}(B)$ ($f(-2, 0) = 1$ y $f(2, 0) = 9$), concluimos que los valores máximo y mínimo de f sobre B son 0 y 9, y que estos valores se alcanzan en los puntos $(-1, 0)$ y $(2, 0)$, respectivamente (ver figura 4.15).

2. *Sea $f : \mathbb{R}^2 \to \mathbb{R}$ dada por*

$$f(x, y) = x(x-1)^2 + y^2$$

y $B = \big\{(x, y) \in \mathbb{R}^2 \mid \|(x, y)\| \leq 1\big\}$.

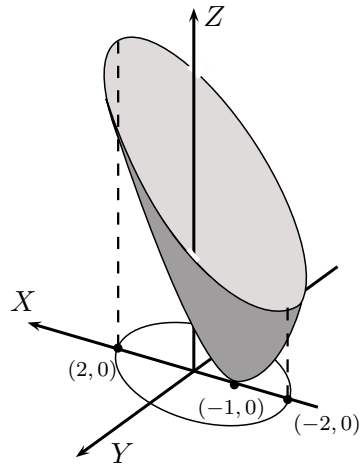


Figura 4.15: Los valores máximo y mínimo de la función $f(x,y) = (x+1)^2+y^2$ sobre el conjunto $B = \{(x,y) \in \mathbb{R}^2 \mid \|(x,y)\| \leq 2\}$ se alcanzan en los puntos $(2,0)$ y $(-1,0)$, respectivamente.

Nuevamente, para determinar los puntos en los que f alcanza un valor máximo o mínimo local (o global) sobre B, procederemos como en el inciso anterior. Primero encontraremos los puntos críticos de f, para lo cual hay que resolver las ecuaciones

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y) &= (x-1)^2 + 2x(x-1)\\ &= (x-1)(3x-1)\\ &= 0\end{aligned}$$

y

$$\begin{aligned}\frac{\partial f}{\partial y}(x,y) &= 2y\\ &= 0\end{aligned}$$

y cuyas soluciones son las parejas $(1,0)$ y $(1/3,0)$, en donde este último punto es el único que queda dentro del interior de B.

Ahora será necesario calcular las segundas derivadas parciales de f, las cuales resultan ser:

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2}(x,y) &= 6x-4\\ \frac{\partial^2 f}{\partial y \partial x}(x,y) &= 0\\ \frac{\partial^2 f}{\partial y^2}(x,y) &= 2,\end{aligned}$$

de modo que para el punto crítico $(1/3,0)$ se tiene que

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2}(1/3,0) &= -2\\ \frac{\partial^2 f}{\partial y \partial x}(1/3,0) &= 0\\ \frac{\partial^2 f}{\partial y^2}(1/3,0) &= 2.\end{aligned}$$

Es decir,

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2}(1/3,0)\frac{\partial^2 f}{\partial y^2}(1/3,0) - \left(\frac{\partial^2 f}{\partial y \partial x}(1/3,0)\right)^2 &= -4\\ &< 0\end{aligned}$$

y por lo tanto, por el problema 50 (o el problema 49), concluimos que f tiene un punto silla en el punto $(1/3,0)$.

Como en el inciso anterior, nos resta localizar los puntos en los que f alcanza sus valores máximo y mínimo sobre la frontera del conjunto B, que ahora es el conjunto

$$\mathrm{Fr}(B) = \left\{(x,y) \in \mathbb{R}^2 \mid \|(x,y)\| = 1\right\}$$

y que, como en el inciso anterior, podemos parametrizar por la función $\gamma : [0,2\pi] \subset \mathbb{R} \to \mathbb{R}^2$ dada por

$$\gamma(t) = (\cos(t), \mathrm{sen}(t)).$$

Como en el caso anterior, definimos $g : [0, 2\pi] \subset \mathbb{R} \to \mathbb{R}$ *como*

$$\begin{aligned} g(t) &= (f \circ \gamma)(t) \\ &= f(\gamma(t)) \\ &= \cos(t)(\cos(t) - 1)^2 + \operatorname{sen}^2(t) \end{aligned}$$

de tal forma que

$$\begin{aligned} g'(t) &= \nabla f(\gamma(t)) \cdot \gamma'(t) \\ &= (\cos(t) - 1)(3\cos(t) - 1)(-\operatorname{sen}(t)) + 2\operatorname{sen}(t)\cos(t) \\ &= \operatorname{sen}(t)\left(2\cos(t) - (\cos(t) - 1)(3\cos(t) - 1)\right) \\ &= \operatorname{sen}(t)\left(-3\cos^2(t) + 6\cos(t) - 1\right). \end{aligned}$$

Dado que las raíces de la ecuación cuadrática $-3x^2 + 6x - 1 = 0$ *son* $1 - \sqrt{6}/3$ *y* $1 + \sqrt{6}/3$, *concluimos que los puntos críticos de* g *en el intervalo abierto* $(0, 2\pi)$ *son* π, t_0, *y* $2\pi - t_0$, *en donde* $t_0 \in (0, \pi)$ *es tal que* $\cos(t_0) = 1 - \sqrt{6}/3$. *Por lo tanto, como* $g(0) = 0$, $g(t_0) = 4\sqrt{6}/9 = g(2\pi - t_0)$, $g(\pi) = -4$ *y* $g(2\pi) = 0$, *concluimos que* f *alcanza su valor máximo sobre la* $\operatorname{Fr}(B)$ *en los puntos*

$$\begin{aligned} \gamma(t_0) &= \left(1 - \sqrt{6}/3, \sqrt{2(\sqrt{6} - 1)/3}\right), \\ \gamma(2\pi - t_0) &= \left(1 - \sqrt{6}/3, -\sqrt{2(\sqrt{6} - 1)/3}\right) \end{aligned}$$

y este máximo es $4\sqrt{6}/9$, *mientras que su valor mínimo lo alcanza en el punto* $\gamma(\pi) = (-1, 0)$ *y este valor es* -4.

En resumen, comparando los valores de f *en los puntos*

$$\begin{gathered} (1, 0), \quad \left(1 - \sqrt{6}/3, \sqrt{2(\sqrt{6} - 1)/3}\right), \\ \left(1 - \sqrt{6}/3, -\sqrt{2(\sqrt{6} - 1)/3}\right), \quad (-1, 0) \end{gathered}$$

(el punto crítico $(1/3, 0)$ *no lo consideramos, puesto que ahí la función* f *tiene un punto silla), concluimos que el valor mínimo de* f *sobre el conjunto* B *es* -4 *y lo alcanza en el punto* $(-1, 0)$*, mientras que su valor máximo es* $4\sqrt{6}/9$ *y lo alcanza en los puntos*

$$\left(1 - \sqrt{6}/3, \sqrt{2(\sqrt{6} - 1)/3}\right) \qquad y \qquad \left(1 - \sqrt{6}/3, -\sqrt{2(\sqrt{6} - 1)/3}\right).$$

4.6.1. Breve comentario sobre formas cuadráticas

En esta breve subsección, justificaremos "intuitivamente" el resultado más conocido relacionado con las formas cuadráticas. En este resultado se establecen (para el caso general en $\mathbb{R}^n$) condiciones necesarias y suficientes para que una forma cuadrática F asociada a una matriz A sea semipositiva (o seminegativa) y no degenerada.

Sin lugar a dudas, las formas cuadráticas del tipo

$$\tilde{F}(x_1, \ldots, x_n) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2 \tag{4.47}$$

son las más simples de analizar, pues es muy fácil comprobar que F es una forma cuadrática semipositiva (seminegativa) y no degenerada si y sólo si $\lambda_i > 0$ ($\lambda_i < 0$) para cada $i \in \{1, \ldots, n\}$ (nótese que $\tilde{F}(\hat{e}_i) = \lambda_i$).

También es fácil comprobar que

$$\tilde{F}(x_1, \ldots, x_n) = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^t,$$

de tal manera que la matriz asociada a $\tilde{F}$ es la matriz diagonal

$$D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}.$$

Un hecho interesante de este tipo de matrices es que si definimos, para cada $k \in \{1, \ldots, n\}$,

$$D_k = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix},$$

entonces

$$\det(D_k) = \lambda_1 \cdots \lambda_k$$

de tal forma que ahora podemos reformular la afirmación anterior de la siguiente manera: $\tilde{F}$ es una forma cuadrática semipositiva (seminegativa) y no degenerada si y sólo si $\det(D_k) > 0$ ($\det(D_k) < 0$ si k es impar y $\det(D_k) > 0$ si k es par) para cada $k \in \{1, \ldots, n\}$.

La buena noticia es que este criterio (por medio de los determinantes de las submatrices D_k) que se cumple para el caso particular de $\tilde{F}$, se sigue cumpliendo en el caso general.

Es decir, si $F : \mathbb{R}^n \to \mathbb{R}$ es una forma cuadrática que tiene asociada la matriz simétrica

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix},$$

se prueba que: F es semipositiva (seminegativa) y no degenerada si y sólo si $\det(A_k) > 0$ ($\det(A_k) < 0$ si k es impar y $\det(A_k) > 0$ si k es par) para cada $k \in \{1, \ldots, n\}$, en donde

$$A_k = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}.$$

Como el lector puede comprobar fácilmente, en la proposición 4.22 se prueba esta afirmación para el caso $n = 2$.

Para reforzar un poco más las ideas anteriores, resulta relevante mencionar un resultado muy importante del Álgebra Lineal (Teorema 6.20 de [2]):

si $A \in M_{n\times n}(\mathbb{R})$ es una matriz simétrica, entonces existe $B \in M_{n\times n}(\mathbb{R})$ una matriz ortonormal tal que BAB^t es una matriz diagonal. Es decir, que

$$BAB^t = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}. \tag{4.48}$$

Como seguramente el lector recordará, las matrices ortonormales son justo las matrices que se obtienen al (y que se usan para) realizar un cambio entre bases ortonormales de $\mathbb{R}^n$. De esta forma, con base en este resultado, podemos afirmar que: si $F : \mathbb{R}^n \to \mathbb{R}$ es una forma cuadrática que tiene asociada la matriz simétrica A (en una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_n\}$), es decir que

$$\begin{aligned} F(\hat{x}) &= F(x_1, \ldots, x_n) \\ &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} A \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^t, \end{aligned}$$

entonces existe otra base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ tal que la matriz asociada a F en esta nueva base es una matriz diagonal.

En efecto, observe que si

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix}$$

es la matriz ortonormal tal que BAB^t es una matriz diagonal, entonces tomando

$$\hat{e}'_i = b_{i1}\hat{e}_1 + \cdots + b_{in}\hat{e}_n$$

para $i \in \{1, \ldots, n\}$, obtenemos que $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ es una base ortonormal de $\mathbb{R}^n$. De acuerdo con la identidad 4.1, si $\hat{x} \in \mathbb{R}^n$ tiene coordenadas $(x'_1, \ldots, x'_n)$ en esta base, entonces sus coordenadas $(x_1, \ldots, x_n)$ en la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$

están dadas por la identidad

$$\begin{aligned}\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} &= \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix} \\ &= \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} B.\end{aligned}$$

De esta manera, la forma cuadrática F expresada en términos de las nuevas coordenadas $x'_1, \ldots, x'_n$ se escribe como

$$\begin{aligned}F(\hat{x}) &= F(x_1, \ldots, x_n) \\ &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} A \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^t \\ &= \left(\begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} B\right) A \left(\begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} B\right)^t \\ &= \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \left(BAB^t\right) \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix}^t \\ &= \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix} \begin{bmatrix} x'_1 & \cdots & x'_n \end{bmatrix}^t \\ &= \lambda_1 \left(x'_1\right)^2 + \cdots + \lambda_n \left(x'_n\right)^2.\end{aligned}$$

Es decir, en la base $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ la función cuadrática F tiene la forma 4.47.

Sin duda que la parte medular de estas últimas ideas está en cómo, dada la matriz A, se encuentra la matriz ortonormal B y los escalares $\lambda_1, \ldots, \lambda_n$ que satisfagan la identidad 4.48. Pero este problema es todo un tema de un curso de Álgebra Lineal.

4.6.2. Máximos y mínimos sobre restricciones

Como se muestra en el ejemplo 4.17, para localizar los puntos en donde una función f alcanza sus valores extremos, no basta con encontrar sus puntos críticos (incluso puede suceder que estos valores extremos no se alcancen en este tipo de puntos, como sucede en el inciso 2 de este mismo ejemplo).

En la mayoría de los casos, la localización de los valores extremos (globales) de una función de $\mathbb{R}^n$ en $\mathbb{R}$, requerirá la localización de valores extremos

sobre conjuntos "más pequeños", cuya característica principal será que son "flacos" (es decir, de interior vacío). En los incisos del ejemplo 4.17 estos conjuntos fueron

$$\left\{(x,y) \in \mathbb{R}^2 \mid \|(x,y)\| = 2\right\} \qquad \text{y} \qquad \left\{(x,y) \in \mathbb{R}^2 \mid \|(x,y)\| = 1\right\},$$

que además de poderse parametrizar por una función de $\mathbb{R}$ en $\mathbb{R}^2$ (propiedad que fue fundamental para encontrar los valores extremos sobre estos conjuntos), también tienen la propiedad de ser los conjuntos de nivel de alguna función de $\mathbb{R}^2$ en $\mathbb{R}$. De hecho, ambos conjuntos son conjuntos de nivel de la función $g(x,y) = x^2 + y^2$ para $c = 4$ y $c = 1$, es decir $N_4(g)$ y $N_1(g)$, respectivamente.

Lo interesante de la observación anterior es que, si ahora nos fijamos en los conjuntos de nivel de la función $f(x,y) = (x+1)^2 + y^2$ (la función del inciso 1 del mencionado ejemplo) para $c = 1$ y $c = 9$ (que son los valores extremos que alcanzó f sobre $N_4(g)$, y que denotamos por $N_1(f)$ y $N_9(f)$, respectivamente), notaremos que estos conjuntos de nivel se intersectan "tangencialmente" con el conjunto de nivel $N_4(g)$ en los puntos $(-2,0)$ y $(2,0)$, que son justo los puntos en donde f alcanza estos valores extremos (ver figura 4.16).

Lo importante de este hecho geométrico es que éste se puede "traducir" en una identidad (es decir, se puede "escribir" en forma "analítica"). En efecto, si ahora recordamos que el gradiente de una función en un punto $\hat{x}_0$ siempre es normal al conjunto de nivel que contiene a este punto, concluimos entonces que los vectores $\nabla f(-2,0)$ y $\nabla g(-2,0)$ deben ser paralelos (es decir, uno debe ser múltiplo del otro), y lo mismo debe suceder para los vectores $\nabla f(2,0)$ y $\nabla g(2,0)$, lo que es muy fácil de verificar, ya que

$$\nabla f(x,y) = (2(x+1), 2y) \qquad \text{y} \qquad \nabla g(x,y) = (2x, 2y)$$

para todo $(x,y) \in \mathbb{R}^2$, de modo que

$$\nabla f(2,0) = (6,0) \qquad \text{y} \qquad \nabla g(2,0) = (4,0),$$

y

$$\nabla f(-2,0) = (-2,0) \qquad \text{y} \qquad \nabla g(-2,0) = (-4,0).$$

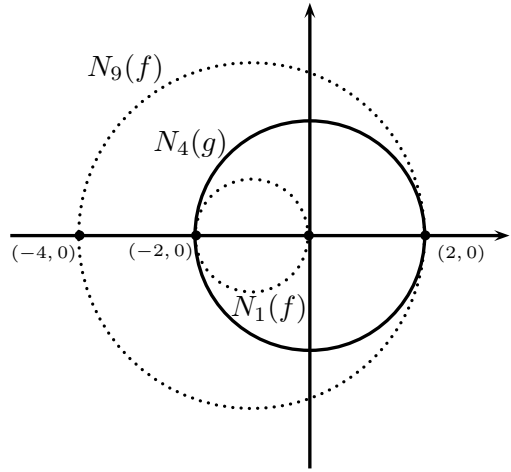


Figura 4.16: Los conjuntos de nivel $N_9(f)$ y $N_1(f)$ (de la función $f(x,y) = (x+1)^2 + y^2$) intersectan tangencialmente al conjunto de nivel $N_4(g)$ (de la función $g(x,y) = x^2 + y^2$) en los puntos $(2,0)$ y $(-2,0)$, respectivamente.

Esto comprueba que

$$\nabla f(2,0) = \frac{3}{2}\nabla g(2,0) \qquad \text{y} \qquad \nabla f(-2,0) = \frac{1}{2}\nabla g(-2,0).$$

La discusión anterior parece sugerir lo siguiente: si tenemos una función f para la cual queremos calcular sus valores extremos sobre un conjunto de nivel de una función g (digamos $N_c(g)$), y localizar los puntos de este conjunto en los cuales alcanza estos valores extremos, es suficiente con encontrar los puntos $\hat{x} \in N_c(g)$ en los cuales se satisface que $\nabla f(\hat{x})$ y $\nabla g(\hat{x})$ son paralelos.

Aprovechando que ya sabemos cuáles son estos puntos para las funciones f y g del inciso 2 del ejemplo 4.17, sigamos el procedimiento descrito en el párrafo anterior y veamos si llegamos a las mismas soluciones. ¡Manos a la obra!

Como $f(x,y) = x(x-1)^2 + y^2$ y $g(x,y) = x^2 + y^2$, se tiene que

$$\begin{aligned}\nabla f(x,y) &= ((x-1)^2 + 2x(x-1), 2y)\\ &= ((x-1)(3x-1), 2y)\end{aligned}$$

y

$$\nabla g(x,y) = (2x, 2y),$$

de modo que nuestro problema es localizar las parejas $(x,y) \in N_1(g)$ para las cuales se satisfaga que

$$\nabla f(x,y) = \lambda \nabla g(x,y)$$

para alguna $\lambda \in \mathbb{R}$.

Lo anterior se traduce en encontrar las parejas $(x,y) \in \mathbb{R}^2$ y los números $\lambda \in \mathbb{R}$ que se satisfagan las siguientes ecuaciones:

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y) &= \lambda \frac{\partial g}{\partial x}(x,y)\\ \frac{\partial f}{\partial y}(x,y) &= \lambda \frac{\partial g}{\partial y}(x,y)\\ g(x,y) &= c\end{aligned}$$

que en nuestro caso se traducen en el sistema de ecuaciones

$$\begin{aligned}(x-1)(3x-1) &= \lambda 2x \qquad (1)\\ 2y &= \lambda 2y \qquad (2)\\ x^2 + y^2 &= 1 \qquad (3)\end{aligned}$$

que en principio se debería poder resolver, pues es un sistema de tres ecuaciones con tres incógnitas (x, y y λ).

Empecemos por observar que, si alguna solución de las ecuaciones anteriores fuera tal que $y \neq 0$, entonces de la ecuación (2) deducimos que $\lambda = 1$, de modo que la ecuación (1) se convierte en la ecuación cuadrática (en la variable x)

$$3x^2 - 6x + 1 = 0,$$

cuyas soluciones son

$$\begin{aligned} x_1 &= \frac{6+\sqrt{6^2-4(3)(1)}}{2(3)} \\ &= 1+\frac{\sqrt{6}}{3} \end{aligned}$$

y

$$\begin{aligned} x_2 &= \frac{6-\sqrt{6^2-4(3)(1)}}{2(3)} \\ &= 1-\frac{\sqrt{6}}{3}. \end{aligned}$$

Si sustituimos estas soluciones en la ecuación (3), concluimos que sólo podemos tomar el segundo valor, obteniendo las siguientes soluciones a nuestro sistema original

$$(x,y) = \left(1-\frac{\sqrt{6}}{3}, \sqrt{\frac{2(\sqrt{6}-1)}{3}}\right) \qquad \text{con} \qquad \lambda = 1$$

y

$$(x,y) = \left(1-\frac{\sqrt{6}}{3}, -\sqrt{\frac{2(\sqrt{6}-1)}{3}}\right) \qquad \text{con} \qquad \lambda = 1.$$

Si ahora suponemos que tenemos una solución de nuestro sistema de ecuaciones con $y = 0$, de la ecuación (3) de este sistema concluimos que $x = 1$ o $x = -1$. Esto, partiendo de la ecuación (1), nos conduce a las soluciones

$$(x,y) = (1,0) \qquad \text{con} \qquad \lambda = 0$$

y

$$(x,y) = (-1,0) \qquad \text{con} \qquad \lambda = 4.$$

Seguramente el lector ya se percató de que con este procedimiento obtuvimos los mismos puntos que obtuvimos en el inciso 2 del ejemplo 4.17.

La buena noticia es que el procedimiento anterior sigue siendo válido en cualquier $\mathbb{R}^n$. Es decir, aseguramos que: si f es una función derivable sobre un conjunto de nivel $N_c(g)$ de una función derivable g, y $\hat{x}_0 \in N_c(g)$ es un punto en el que la función f tiene un máximo o mínimo local (sobre $N_c(g)$), entonces se debe cumplir que $\nabla f(\hat{x}_0)$ y $\nabla g(\hat{x}_0)$ son vectores paralelos, es decir, que existe $\lambda \in \mathbb{R}$ tal que

$$\nabla f(\hat{x}_0) = \lambda \nabla g(\hat{x}_0).$$

La importancia de la afirmación anterior es que, si deseamos localizar los puntos en los que una función f alcanza sus valores extremos (incluso locales) sobre un conjunto de nivel $N_c(g)$ de una función derivable g, basta con encontrar los puntos de este conjunto para los cuales se satisface que los gradientes de f y de g son paralelos. Ilustraremos la afirmación anterior con un problema geométrico (en $\mathbb{R}^3$) que sin duda el lector conoce (y sabe la respuesta): dado un plano P cuya ecuación está dada por

$$Ax + By + Cz + D = 0$$

(con $A^2 + B^2 + C^2 > 0$) y un punto $\hat{x}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$, calcular la distancia del punto $\hat{x}_0$ al plano P.

Como sabemos, la distancia del punto $\hat{x}_0$ al plano P se obtiene como la mínima distancia entre el punto $\hat{x}_0$ y los puntos del plano P. Es decir, podemos plantear nuestro problema de la siguiente manera: consideramos la función d que nos da la distancia entre el punto $\hat{x}_0$ y cualquier otro punto $\hat{x} = (x, y, z) \in \mathbb{R}^3$, es decir

$$d(x, y, z) = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}.$$

Nos planteamos el problema de localizar el mínimo valor de la función d sobre el conjunto $N_0(g)$, en donde g la definimos como

$$g(x, y, z) = Ax + By + Cz + D.$$

Para simplificar los cálculos, dado que el punto en $N_0(g)$ en el que se minimiza la distacia con $\hat{x}_0$ es el mismo en el que se minimiza la distacia al

cuadrado, trabajaremos con la función

$$\begin{aligned} f(x,y,z) &= d^2(x,y,z) \\ &= (x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2 \end{aligned}$$

(esta es una simplificación que el lector debería de tener siempre muy presente).

Dado que

$$\nabla f(x,y,z) = (2(x-x_0), 2(y-y_0), 2(z-z_0))$$

y

$$\nabla g(x,y,z) = (A,B,C)$$

nuestro problema se reduce a resolver el sistema de ecuaciones

$$\begin{aligned} 2(x-x_0) &= \lambda A \\ 2(y-y_0) &= \lambda B \\ 2(z-z_0) &= \lambda C \\ Ax + By + Cz + D &= 0. \end{aligned}$$

De las primeras tres ecuaciones concluimos que

$$\begin{aligned} x &= \frac{1}{2}\lambda A + x_0 \\ y &= \frac{1}{2}\lambda B + y_0 \\ z &= \frac{1}{2}\lambda C + z_0. \end{aligned}$$

Si sustituimos estos valores en la cuarta ecuación, obtenemos que

$$A\left(\frac{1}{2}\lambda A + x_0\right) + B\left(\frac{1}{2}\lambda B + y_0\right) + C\left(\frac{1}{2}\lambda C + z_0\right) + D = 0,$$

es decir, que

$$\lambda = -\frac{2\left(Ax_0 + By_0 + Cz_0 + D\right)}{A^2 + B^2 + C^2}. \tag{4.49}$$

Por lo tanto, sustituyendo λ en los valores despejados de x, y y z, el punto $\hat{x}_p$ dado por la terna

$$\begin{aligned}\hat{x}_p &= \left(x_0 - A\frac{Ax_0+By_0+Cz_0+D}{A^2+B^2+C^2}, y_0 - B\frac{Ax_0+By_0+Cz_0+D}{A^2+B^2+C^2},\right.\\ &\qquad \left. z_0 - C\frac{Ax_0+By_0+Cz_0+D}{A^2+B^2+C^2}\right)\\ &= \hat{x}_0 - \frac{Ax_0+By_0+Cz_0+D}{A^2+B^2+C^2}(A,B,C)\end{aligned}$$

es el único punto del plano P para el cual se satisface que

$$\nabla f(\hat{x}_p) = \lambda \nabla g(\hat{x}_p),$$

en donde λ está dada por 4.49 (nótese que, si $\hat{x}_0 \in P$, entonces $\hat{x}_p = \hat{x}_0$, ¡como era de esperarse!), de modo que la distacia del punto $\hat{x}_0$ al plano P estará dada por

$$\begin{aligned}d(\hat{x}_0, P) &= \sqrt{f(\hat{x}_p)}\\ &= \|\hat{x}_p - \hat{x}_0\|\\ &= \left\| -\frac{Ax_0+By_0+Cz_0+D}{A^2+B^2+C^2}(A,B,C)\right\|\\ &= \frac{|Ax_0+By_0+Cz_0+D|}{\|(A,B,C)\|}\\ &= \frac{|Ax_0+By_0+Cz_0+D|}{\sqrt{A^2+B^2+C^2}},\end{aligned}$$

lo que sin duda el lector recordará muy bien de sus cursos de Geometría Analítica.

El ejemplo anterior no sólo refuerza nuestra sospecha de que, si una función f alcanza un valor extremo sobre un conjunto de nivel de otra función g en un punto $\hat{x}_0$ de este conjunto, entonces los vectores gradiente de f y g en $\hat{x}_0$ deben ser paralelos, sino que nos permite plantear un problema

que nos conducirá a formular un resultado más general que el descrito en el párrafo anterior.

El problema ahora es el siguiente:

1. dada una recta l determinada por la intersección de los planos P_1 y P_2 cuya ecuación está dada por

$$\begin{aligned} A_1x + B_1y + C_1z + D_1 &= 0 \qquad \text{y} \\ A_2x + B_2y + C_2z + D_2 &= 0 \end{aligned}$$

respectivamente, (con (A_1, B_1, C_1) y (A_2, B_2, C_2) vectores no paralelos (en particular diferentes de $\hat{0}$) para que sí determinen una recta), y

2. dado un punto $\hat{x}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$,

calcular la distacia del punto $\hat{x}_0$ a la recta l.

Como en el ejemplo anterior, usaremos la función que nos calcula la distacia al cuadrado entre el punto $\hat{x}_0$ y un punto de la recta l. Nuestra tarea será localizar el punto de l en el que esta función alcanza su valor mínimo (con respecto a los puntos de la recta l). Es decir, tomaremos la función

$$\begin{aligned} f(x, y, z) &= d^2(x, y, z) \\ &= (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \end{aligned}$$

y nuestro objetivo será localizar un punto $\hat{x}_l$ del conjunto

$$S = \big\{(x, y, z) \in \mathbb{R}^3 \mid g_1(x, y, z) = 0, g_2(x, y, z) = 0\big\}$$

tal que f alcance su valor mínimo (con respecto a los puntos de S) en $\hat{x}_l$, en donde g_1 y g_2 están dadas por

$$\begin{aligned} g_1(x, y, z) &= A_1x + B_1y + C_1z + D_1 \\ g_2(x, y, z) &= A_2x + B_2y + C_2z + D_2. \end{aligned}$$

Desde un punto de vista geométrico, si $\hat{x}_l$ es el punto de la recta l que está más cerca del punto $\hat{x}_0$, esto significa que $\hat{x}_l$ es el único punto de esta recta que pertenece al conjunto de nivel $N_c(f)$ de f, en donde

$$\begin{aligned} c &= f(\hat{x}_l) \\ &= \|\hat{x}_l - \hat{x}_0\|^2 \end{aligned}$$

y $N_c(f)$ es la esfera de radio $\sqrt{c}$ con centro en $\hat{x}_0$. Dicho de otra forma, la recta l es "tangente" a la esfera $N_c(f)$ en el punto $\hat{x}_l$. Nótese ahora que, si tomamos el vector que va de $\hat{x}_0$ a $\hat{x}_l$, que llamaremos $\hat{v}$, este debe ser perpendicular a la recta l, y si por otra parte recordamos que los vectores (A_1, B_1, C_1) y (A_2, B_2, C_2) son normales a los planos P_1 y P_2 (que determinan a la recta l y que por lo tanto la contienen), tendremos que dichos vectores también deberán ser perpendiculares a l (ver figura 4.17).

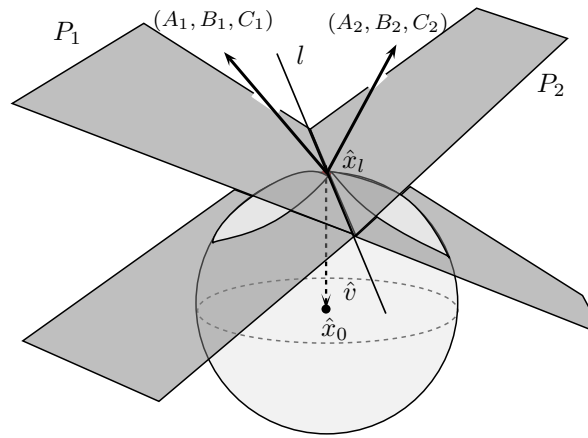


Figura 4.17: Dado que la recta l, que es la intersección de los planos P_1 y P_2, es tangente en el punto $\hat{x}_l$ a la esfera con centro en $\hat{x}_0$, los vectores $\hat{v} = \hat{x}_l - \hat{x}_0$, (A_1, B_1, C_1) y (A_2, B_2, C_2) (los vectores normales a P_1 y P_2, respectivamente) son perpendiculares a l, y por lo mismo pertenecen a un mismo plano.

Del razonamiento anterior concluimos que, como los vectores $\hat{v}$, (A_1, B_1, C_1) y (A_2, B_2, C_2) son perpendiculares a la misma recta l, y los dos últi-

mos son linealmente independientes, se debe cumplir que el vector $\hat{v}$ se puede escribir como una combinación lineal de los vectores (A_1, B_1, C_1) y (A_2, B_2, C_2).

Finalmente, si ahora observamos:

1. que el vector (A_1, B_1, C_1) coincide con ser el vector gradiente de g_1 evaluado en cualquier punto del plano P_1 (en particular en $\hat{x}_l$),

2. que lo mismo sucede con el vector (A_2, B_2, C_2) y la función g_2, y

3. que los vectores $\nabla f(\hat{x}_l)$ y $\hat{v}$ son paralelos, pues ambos son normales a la esfera $N_c(f)$,

concluimos que el vector $\nabla f(\hat{x}_l)$ se debe poder expresar como una combinación lineal de los vectores $\nabla g_1(\hat{x}_l)$ y $\nabla g_2(\hat{x}_l)$. Es decir, deben existir $\lambda_1, \lambda_2 \in \mathbb{R}$ tales que

$$\nabla f(\hat{x}_l) = \lambda_1 \nabla g_1(\hat{x}_l) + \lambda_2 \nabla g_2(\hat{x}_l).$$

Para aprovechar la discusión anterior, formularemos un resultado mas general que seguramente el lector ya está intuyendo: sea $S \subset \mathbb{R}^3$ la intersección de los conjuntos de nivel (que sin perdida de generalidad supondremos que son los conjuntos de nivel 0) de dos funciones derivables g_1 y g_2, es decir que

$$S = \left\{(x, y, z) \in \mathbb{R}^3 \mid g_1(x, y, z) = 0, g_2(x, y, z) = 0\right\}.$$

Si $\hat{x}_0 \in S$ es un punto tal que:

1. los vectores $\nabla g_1(\hat{x}_0)$ y $\nabla g_2(\hat{x}_0)$ son linealmente independientes, y

2. una función f (derivable) alcanza un valor extremo (incluso local) en $\hat{x}_0$ sobre (o con respecto a) S,

entonces se debe cumplir que el vector $\nabla f(\hat{x}_0)$ se puede expresar como una combinación lineal de los vectores $\nabla g_1(\hat{x}_0)$ y $\nabla g_2(\hat{x}_0)$, es decir, deben existir $\lambda_1, \lambda_2 \in \mathbb{R}$ tales que

$$\nabla f(\hat{x}_0) = \lambda_1 \nabla g_1(\hat{x}_0) + \lambda_2 \nabla g_2(\hat{x}_0).$$

Si el lector ya está covencido de que el resultado anterior debe ser cierto, seguramente estará de acuerdo en que su siguiente generalización (para funciones de $\mathbb{R}^n$ en $\mathbb{R}$) también debe ser cierta: si $g_1, \ldots, g_m : U \subset \mathbb{R}^n \to \mathbb{R}$ son m funciones de clase C^1 en U, con $m < n$, $S \subset \mathbb{R}^n$ es el conjunto dado por

$$S = \{\hat{x} \in U \subset \mathbb{R}^n \mid g_1(\hat{x}) = 0, \ldots, g_m(\hat{x}) = 0\}$$

y $\hat{x}_0 \in S$ es un punto tal que:

1. los vectores $\nabla g_1(\hat{x}_0), \ldots, \nabla g_m(\hat{x}_0)$ son linealmente independientes, y
2. una función $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^1 en U alcanza un valor extremo (incluso local) en $\hat{x}_0$ sobre (o con respecto a) S,

entonces se debe cumplir que el vector $\nabla f(\hat{x}_0)$ se puede expresar como una combinación lineal de los vectores $\nabla g_1(\hat{x}_0), \ldots, \nabla g_m(\hat{x}_0)$. Es decir, deben existir $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ tales que

$$\nabla f(\hat{x}_0) = \lambda_1 \nabla g_1(\hat{x}_0) + \cdots + \lambda_m \nabla g_m(\hat{x}_0).$$

Todo el trabajo que nos hemos tomado a lo largo de esta sección ha sido con el único objetivo de llegar al resultado anterior, el cual es conocido como el *Teorema de los multiplicadores de Lagrange*[4]. Antes de formularlo de manera más formal, conviene hacer algunas observaciones, empezando por su nombre: el término "multiplicadores" se refiere a los escalares $\lambda_1, \ldots, \lambda_m$. Al conjunto S se le conoce con el nombre de "*conjunto de restricciones*" y la condición de que m sea menor a n tiene que ver con los siguientes dos hechos: uno, que sólo si m es menor o igual a n se puede cumplir que m vectores en $\mathbb{R}^n$ sean linealmente independientes, y dos, si $m = n$ lo más probable es que el conjunto S sea vacío, o contega a lo más un número finito de puntos.

[4]Joseph-Louis Lagrange, bautizado como Giuseppe Lodovico Lagrangia, también llamado Giuseppe Luigi Lagrangia o Lagrange (Turín, 25 de enero de 1736 - París, 10 de abril de 1813), fue un físico, matemático y astrónomo italiano que después vivió en Prusia y Francia. Lagrange trabajó para Federico II de Prusia, en Berlín, durante veinte años. Lagrange demostró el teorema del valor medio, desarrolló la mecánica Lagrangiana y tuvo una importante contribución en astronomía. (fuente: Wikipedia).

Finalmente, la condición de que las funciones $g_1, \ldots, g_m$ y f sean de clase C^1 en U está directamente relacionada con las "herramientas" que usaremos en su prueba, que por cierto, daremos hasta el siguiente capítulo, pues mucha de esta "herramienta" será desarrollada hasta entonces.

La versión que probaremos en este texto del teorema de los multiplicadores de Lagrange es la siguiente.

Teorema 4.3 (de los multiplicadores de Lagrange) *Sean:*

1. $g_1, \ldots, g_m : U \subset \mathbb{R}^n \to \mathbb{R}$ *funciones de clase* C^1 *en* U*, con* $m < n$*,*
2. $S \subset \mathbb{R}^n$ *el conjunto dado por*

$$S = \{\hat{x} \in U \subset \mathbb{R}^n \mid g_1(\hat{x}) = 0, \ldots, g_m(\hat{x}) = 0\}$$

3. $\hat{x}_0 \in S$ *tal que* $\nabla g_1(\hat{x}_0), \ldots, \nabla g_m(\hat{x}_0)$ *son linealmente independientes.*

Si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ *es una función de clase* C^1 *en* U *tal que* f *tiene un máximo o mínimo (global o sólo local) en* $\hat{x}_0$ *sobre* S*, entonces existen* $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ *tales que*

$$\nabla f(\hat{x}_0) = \lambda_1 \nabla g_1(\hat{x}_0) + \cdots + \lambda_m \nabla g_m(\hat{x}_0).$$

Dado que la prueba de este teorema tendrá que esperar hasta el siguiente capítulo, por ahora sólo haremos una importante observación con relación a la tercera hipótesis de su enunciado. Esta observación se relaciona con el hecho de que un mismo conjunto de restricciones S se puede obtener usando diferentes colecciones de funciones $g_1, \ldots, g_m$, y no en todos los casos estas colecciones satisfacen la condición del inciso 3, la cual es fundamental para la validez del teorema. En el ejemplo siguiente ilustramos esta situación.

Ejemplo 4.18 *Sea* $S = \{(t, 0, 0) \in \mathbb{R}^3 \mid t \in \mathbb{R}\}$ *y*

$$\begin{aligned} f(x, y, z) &= \|(x, y, z) - (0, -1, -1)\|^2 \\ &= x^2 + (y + 1)^2 + (z + 1)^2 \end{aligned}$$

(el cuadrado de la distancia entre el punto (x, y, z) y el punto $(0, -1, -1)$).

Seguramente el lector estará de acuerdo en que $\hat{0} = (0, 0, 0)$ es el punto en el cual la función f alcanza su mínimo valor (global) sobre el conjunto S y que en este punto se tiene que

$$\nabla f(\hat{0}) = (0, 2, 2).$$

Primero observemos que, si $g_1(x, y, z) = z$ y $g_2(x, y, z) = z+y^2$, entonces el conjunto S coincide con ser

$$S = \left\{(x, y, z) \in \mathbb{R}^3 \mid g_1(x, y, z) = 0, g_2(x, y, z) = 0\right\}$$

y si $g_3(x, y, z) = y$ también se tiene que

$$S = \left\{(x, y, z) \in \mathbb{R}^3 \mid g_1(x, y, z) = 0, g_3(x, y, z) = 0\right\}.$$

En el primer caso se tiene que $\nabla g_1(\hat{0}) = (0, 0, 1) = \nabla g_2(\hat{0})$, es decir, no son linealmente independientes, y el vector $\nabla f(\hat{0}) = (0, 2, 2)$ no se puede escribir como combinación lineal de ellos (ver figura 4.18). En el segundo caso se tiene que $\nabla g_1(\hat{0}) = (0, 0, 1)$ y $\nabla g_3(\hat{0}) = (0, 1, 0)$, los cuales sí son linealmente independientes y se verifica que

$$\nabla f(\hat{0}) = 2\nabla g_1(\hat{0}) + 2\nabla g_3(\hat{0}),$$

que es lo que asegura el teorema de los multiplicadores de Lagrange (ver figura 4.19).

A continuación mostraremos cómo se usa el teorema 4.3 resolviendo el problema con el cual lo motivamos: encontrar la distancia de un punto $\hat{x}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$ a una recta l, la cual está determinada como la intersección de dos planos.

Desafortunadamente, resolver el caso general planteado inicialmente implica realizar largas y tediosas manipulaciones algebraicas, razón por la cual nos limitaremos a resolver el problema para un punto y una recta específicos.

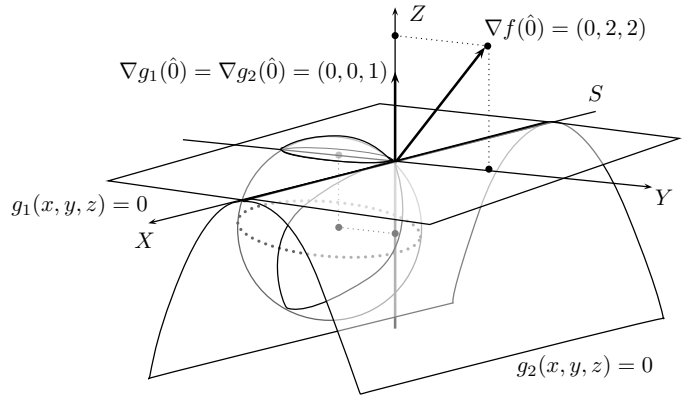


Figura 4.18: En el ejemplo 4.18, los vectores $\nabla g_1(\hat{0}) = \nabla g_2(\hat{0}) = (0,0,1)$ no son linealmente independientes y el vector $\nabla f(\hat{0}) = (0,2,2)$ no se puede escribir como combinación lineal de ellos.

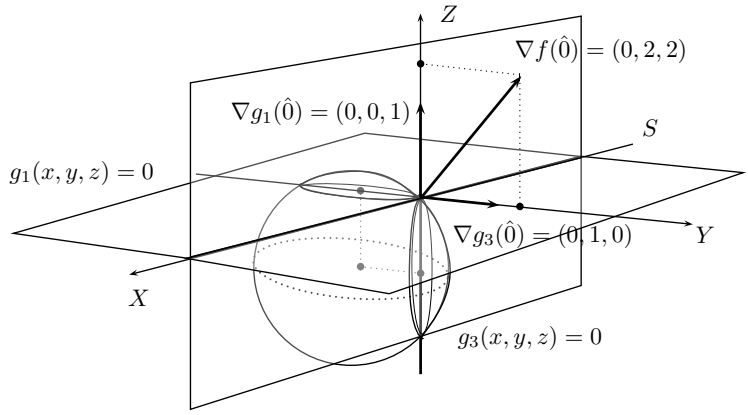


Figura 4.19: En el ejemplo 4.18, los vectores $\nabla g_1(\hat{0}) = (0,0,1)$ y $\nabla g_3(\hat{0}) = (0,1,0)$ sí son linealmente independientes y el vector $\nabla f(\hat{0}) = (0,2,2)$ sí se puede escribir como combinación lineal de ellos, como asegura el teorema de los multiplicadores de Lagrange.

Ejemplo 4.19 *Tomemos el punto* $(1,1,1) \in \mathbb{R}^3$ *y la recta determinada por la intersección de los planos*

$$\begin{aligned} x + y - 1 &= 0 \\ y - z + 1 &= 0, \end{aligned}$$

y calculemos el valor mínimo de la función

$$f(x,y,z) = (x-1)^2 + (y-1)^2 + (z-1)^2$$

sobre el conjunto $S \subset \mathbb{R}^3$ *dado por*

$$S = \left\{ (x,y,z) \in \mathbb{R}^3 \mid g_1(x,y,z) = 0, g_2(x,y,z) = 0 \right\},$$

en donde $g_1(x,y,z) = x + y - 1$ *y* $g_2(x,y,z) = y - z + 1$.

De acuerdo con el teorema de los multiplicadores de Lagrange, si $\hat{x}_0 = (x_0, y_0, z_0) \in S$ *es el punto en el que* f *alcanza su valor mínimo (sobre* S*), deben existir* $\lambda_1, \lambda_2 \in \mathbb{R}$ *tales que*

$$\nabla f(\hat{x}_0) = \lambda_1 \nabla g_1(\hat{x}_0) + \lambda_2 \nabla g_2(\hat{x}_0).$$

Esto nos lleva a resolver el sistema de ecuaciones

$$\begin{aligned} \frac{\partial f}{\partial x}(\hat{x}_0) &= \lambda_1 \frac{\partial g_1}{\partial x}(\hat{x}_0) + \lambda_2 \frac{\partial g_2}{\partial x}(\hat{x}_0) \\ \frac{\partial f}{\partial y}(\hat{x}_0) &= \lambda_1 \frac{\partial g_1}{\partial y}(\hat{x}_0) + \lambda_2 \frac{\partial g_2}{\partial y}(\hat{x}_0) \\ \frac{\partial f}{\partial z}(\hat{x}_0) &= \lambda_1 \frac{\partial g_1}{\partial z}(\hat{x}_0) + \lambda_2 \frac{\partial g_2}{\partial z}(\hat{x}_0) \\ g_1(\hat{x}_0) &= 0 \\ g_2(\hat{x}_0) &= 0 \end{aligned}$$

que en el caso particular que estamos analizando, se traduce en

$$\begin{aligned} 2(x_0 - 1) &= \lambda_1 \\ 2(y_0 - 1) &= \lambda_1 + \lambda_2 \\ 2(z_0 - 1) &= -\lambda_2 \\ x_0 + y_0 - 1 &= 0 \\ y_0 - z_0 + 1 &= 0. \end{aligned}$$

Para resolver este sistema, de las primeras tres ecuaciones despejamos a x_0, y_0 y z_0 en términos de λ_1 y λ_2 y los sustituimos en las últimas dos ecuaciones, con lo que obtenemos el siguiente sistema en las incógnitas λ_1 y λ_2:

$$\begin{aligned} 2\lambda_1 + \lambda_2 &= -2 \\ \lambda_1 + 2\lambda_2 &= -2, \end{aligned}$$

que tiene como solución $\lambda_1 = -2/3 = \lambda_2$. Sustituyendo estos valores en las primeras tres ecuaciones del sistema original, concluimos que

$$(x_0, y_0, z_0) = \frac{1}{3}(2, 1, 4),$$

de manera que la distancia que se estaba buscando es

$$\begin{aligned} \sqrt{f\left(\frac{1}{3}(2,1,-4)\right)} &= \sqrt{\left(\frac{2}{3}-1\right)^2 + \left(\frac{1}{3}-1\right)^2 + \left(\frac{4}{3}-1\right)^2} \\ &= \frac{\sqrt{6}}{3}. \end{aligned}$$

Si el lector resolvió problemas de este estilo en sus cursos de Geometría Analítica, seguramente notó que en general los cálculos que acabamos de hacer son más sencillos, además de que se obtuvo explícitamente cuál es el punto de la recta que está más cercano al punto $\hat{x}_0$, lo que no se suele hacer por los métodos desarrollados en esos cursos.

Finalmente, daremos otro ejemplo de cómo usar el Teorema de los Multiplicadores de Lagrange para demostrar algunas desigualdades importantes. Aun cuando en el capítulo 1 dimos una prueba de la desigualdad de Cauchy-Schwarz, en el siguiente ejemplo daremos otra prueba usando este teorema.

Ejemplo 4.20 *Como sabemos, la desigualdad de Cauchy-Schwarz establece que*

$$|\hat{x} \cdot \hat{y}| \leq \|\hat{x}\| \, \|\hat{y}\|$$

para todos $\hat{x}, \hat{y} \in \mathbb{R}^n$.

Dado que, si $\hat{x} = \hat{0}$ *o* $\hat{y} = \hat{0}$ *es claro que esta desigualdad se satisface, la probaremos para el caso en que* $\hat{x} \neq \hat{0}$ *y* $\hat{y} \neq \hat{0}$. *Bajo este supuesto podemos dividir la desigualdad anterior por* $\|\hat{x}\| \|\hat{y}\|$ *y nuestro problema se reduce a probar que*

$$\left| \frac{\hat{x}}{\|\hat{x}\|} \cdot \frac{\hat{y}}{\|\hat{y}\|} \right| \leq 1.$$

Como

$$\left\| \frac{\hat{x}}{\|\hat{x}\|} \right\| = \frac{1}{\|\hat{x}\|} \|\hat{x}\| = 1 = \frac{1}{\|\hat{y}\|} \|\hat{y}\| = \left\| \frac{\hat{y}}{\|\hat{y}\|} \right\|$$

lo que debemos probar es que

$$|\hat{x} \cdot \hat{y}| \leq 1$$

para todos $\hat{x}, \hat{y} \in \mathbb{R}^n$ *tales que* $\|\hat{x}\| = 1 = \|\hat{y}\|$.

Con base en lo anterior, definimos $f, g_1, g_2 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *como*

$$\begin{aligned} f(x_1, \ldots, x_n, y_1, \ldots, y_n) &= x_1 y_1 + \cdots + x_n y_n \\ g_1(x_1, \ldots, x_n, y_1, \ldots, y_n) &= x_1^2 + \cdots + x_n^2 - 1 \\ g_2(x_1, \ldots, x_n, y_1, \ldots, y_n) &= y_1^2 + \cdots + y_n^2 - 1 \end{aligned}$$

y

$$S = \{(x_1, \ldots, x_n, y_1, \ldots, y_n) \in \mathbb{R}^n \times \mathbb{R}^n \,|\, g_i(x_1, \ldots, x_n, y_1, \ldots, y_n) = 0, i = 1, 2\}$$

de tal forma que ahora nuestro objetivo será mostrar, apoyados en el Teorema de los multiplicadores de Lagrange, que los valores mínimo y máximo de f *sobre el conjunto* S *son* 1 *y* -1, *respectivamente, los cuales deben de existir puesto que* f *es continua en* $\mathbb{R}^n$ *y* S *es un conjunto cerrado y acotado (es decir, compacto).*

Si $(\hat{x}_0, \hat{y}_0) = \left(x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) \in S$ *es un punto en donde* f *alcanza uno de estos valores extremos, por dicho teorema sabemos que existen* $\lambda_1, \lambda_2 \in \mathbb{R}$ *tales que*

$$\nabla f(\hat{x}_0, \hat{y}_0) = \lambda_1 \nabla g_1(\hat{x}_0, \hat{y}_0) + \lambda_2 \nabla g_2 f(\hat{x}_0, \hat{y}_0),$$

es decir, que

$$\begin{aligned} y_i^{(0)} &= 2\lambda_1 x_i^{(0)} \\ x_i^{(0)} &= 2\lambda_2 y_i^{(0)} \end{aligned} \tag{4.50}$$

para cada $i \in \{1, \ldots, n\}$.

Si a la primera de estas identidades la multiplicamos por $x_i^{(0)}$*, sumamos sobre el índice* i*, y usamos el hecho de que*

$$\begin{aligned} \left(x_1^{(0)}\right)^2 + \cdots + \left(x_n^{(0)}\right)^2 - 1 &= g_1\left(x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) \\ &= 0, \end{aligned}$$

obtenemos que

$$\begin{aligned} x_1^{(0)} y_1^{(0)} + \cdots + x_n^{(0)} y_n^{(0)} &= 2\lambda_1 \left(\left(x_1^{(0)}\right)^2 + \cdots + \left(x_n^{(0)}\right)^2\right) \\ &= 2\lambda_1. \end{aligned}$$

Análogamente, si a la segunda identidad de 4.50 la multiplicamos por $y_i^{(0)}$*, sumamos sobre el índice* i*, y ahora usamos el hecho de que*

$$\begin{aligned} \left(y_1^{(0)}\right)^2 + \cdots + \left(y_n^{(0)}\right)^2 - 1 &= g_2\left(x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) \\ &= 0, \end{aligned}$$

obtenemos que

$$\begin{aligned} x_1^{(0)} y_1^{(0)} + \cdots + x_n^{(0)} y_n^{(0)} &= 2\lambda_2 \left(\left(y_1^{(0)}\right)^2 + \cdots + \left(y_n^{(0)}\right)^2\right) \\ &= 2\lambda_2. \end{aligned}$$

Como resultado de estos dos procedimientos, concluimos que $\lambda_1 = \lambda_2$*, de tal forma que si llamamos* $\lambda = \lambda_1 = \lambda_2$*, y sustituimos este valor en las identidades 4.50, obtenemos que*

$$\begin{aligned} y_i^{(0)} &= 2\lambda x_i^{(0)} = 2\lambda\left(2\lambda y_i^{(0)}\right) = 4\lambda^2 y_i^{(0)} \\ x_i^{(0)} &= 2\lambda y_i^{(0)} = 2\lambda\left(2\lambda x_i^{(0)}\right) = 4\lambda^2 x_i^{(0)} \end{aligned}$$

para cada $i \in \{1, \ldots, n\}$.

Dado que $y_i^{(0)}$ *o* $x_i^{(0)}$ *debe ser distinto de* 0 *para alguna* $i \in \{1, \ldots, n\}$, *se tiene que*

$$4\lambda^2 = 1,$$

y por lo tanto que

$$\lambda = \pm\frac{1}{2}.$$

Si sustituimos este valor de λ *en cualquiera de las identidades 4.50, se tiene que*

$$y_i^{(0)} = \pm x_i^{(0)}$$

o

$$x_i^{(0)} = \pm y_i^{(0)}.$$

Nótese que en cualquiera de estos dos casos, obtenemos que

$$\begin{aligned} f\left(x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) &= f\left(x_1^{(0)}, \ldots, x_n^{(0)}, \pm x_1^{(0)}, \ldots, \pm x_n^{(0)}\right) \\ &= x_1^{(0)}\left(\pm x_1^{(0)}\right) + \cdots + x_n^{(0)}\left(\pm x_n^{(0)}\right) \\ &= \pm\left(\left(x_1^{(0)}\right)^2 + \cdots + \left(x_n^{(0)}\right)^2\right) \\ &= \pm 1 \end{aligned}$$

o

$$\begin{aligned} f\left(x_1^{(0)}, \ldots, x_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) &= f\left(\pm y_1^{(0)}, \ldots, \pm y_n^{(0)}, y_1^{(0)}, \ldots, y_n^{(0)}\right) \\ &= y_1^{(0)}\left(\pm y_1^{(0)}\right) + \cdots + y_n^{(0)}\left(\pm y_n^{(0)}\right) \\ &= \pm\left(\left(y_1^{(0)}\right)^2 + \cdots + \left(y_n^{(0)}\right)^2\right) \\ &= \pm 1, \end{aligned}$$

de donde concluimos que el valor mínimo de f *sobre el conjunto* S *es* -1, *y el máximo es* 1. *Es decir que*

$$|\hat{x} \cdot \hat{y}| = |f(\hat{x}, \hat{y})| \leq 1$$

para todos $\hat{x}, \hat{y} \in \mathbb{R}^n$ *tales que* $\|\hat{x}\| = 1 = \|\hat{y}\|$*, que es lo que deseábamos demostrar.*

4.7. Problemas

1. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Pruebe que, si la derivada direccional $D_{\hat{u}} f(\hat{x}_0)$ existe, entonces la derivada direccional $D_{-\hat{u}} f(\hat{x}_0)$ también existe y además

$$D_{-\hat{u}} f(\hat{x}_0) = -D_{\hat{u}} f(\hat{x}_0).$$

2. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $\hat{u} \in \mathbb{R}^n$ tal que $\|\hat{u}\| = 1$. Si $r > 0$ es tal que $B_r(\hat{x}_0) \subset U$, definimos $g : (-r, r) \subset \mathbb{R} \to \mathbb{R}$ como $g(x) = f(\hat{x}_0 + x\hat{u})$. Pruebe que, si $D_{\hat{u}} f(\hat{x})$ existe para toda $\hat{x} \in B_r(\hat{x}_0)$, entonces g es derivable para toda $x \in (-r, r)$ y además

$$g'(x) = D_{\hat{u}} f(\hat{x}_0 + x\hat{u}) \cdot \hat{u}.$$

3. Pruebe las proposiciones 4.1 y 4.2.

4. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ tal que $f(x, y) = f(y, x)$ para todo $(x, y) \in \mathbb{R}^2$. Pruebe que $\frac{\partial f}{\partial x}(a, b) = \frac{\partial f}{\partial y}(b, a)$ para todo $(a, b) \in \mathbb{R}^2$.

5. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ tal que U es un conjunto abierto e $i \in \{1, \ldots, n\}$ fija. Pruebe que, si para todo par de puntos $\hat{x}, \hat{y} \in U$ tales que $\hat{x} - \hat{y} = \lambda \hat{e}_i$ para alguna $\lambda \in \mathbb{R}$ se tiene que $f(\hat{x}) = f(\hat{y})$ (es decir, que f no depende de la variable x_i), entonces $\frac{\partial f}{\partial x_i}$ existe para toda $\hat{x} \in U$ y además

$$\frac{\partial f}{\partial x_i}(\hat{x}) = 0.$$

   ¿Se cumple lo recíproco? ¿Es necesaria otra hipótesis sobre el conjunto U? Pruebe sus respuestas.

6. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ tal que f es constante sobre el conjunto abierto U. Pruebe que f es derivable para toda $\hat{x} \in U$ y que además $Df(\hat{x})$

es la función lineal constante 0 ($Df(\hat{x}) \equiv 0$) para toda $\hat{x} \in U$. ¿Se cumple lo recíproco? ¿Es necesaria otra hipótesis sobre el conjunto U? Pruebe sus respuestas.

7. Si $L : \mathbb{R}^n \to \mathbb{R}$ es una función lineal, pruebe que L es derivable para toda $\hat{x} \in \mathbb{R}^n$ y que $DL(\hat{x}) = L$.

8. Dé una estimación para las siguientes cantidades. Justifique su respuesta.

   i) $(0.99e^{0.02})^8$.

   ii) $(0.99)^3 + (2.01)^3 - 6(0.99)(2.01)$.

   iii) $((4.01)^2 + (3.98)^2 + (2.02)^2)^{1/2}$.

9. El capitán Nemo está en problemas cerca de la parte soleada de Mercurio. La temperatura del casco de su nave, cuando se localiza en el punto (x, y, z), está dada por la función $T(x, y, z) = e^{-x-2y^2-3z^2}$, donde x, y y z están medidos en metros. Si la nave se encuentra en el punto $(1, 1, 1)$:

   a) ¿en qué dirección debe mover la nave para que la temperatura decrezca más rápidamente?

   b) si la nave viaja a e^8 metros por segundo, ¿con qué rapidez decrecerá la temperatura si se mueve en la dirección del inciso anterior?

   c) desafortunadamente, el metal del casco se fracturará si este se enfría a una razón mayor de $\sqrt{14e^2}$ grados por segundo. Describa el conjunto de posibles direcciones en las que se puede mover la nave para que la temperatura decrezca a una razón menor que ésta.

10. Sea $f : I \subset \mathbb{R} \to \mathbb{R}$ derivable en $x_0 \in I$, con I un intervalo abierto. Definimos $h : I \times \mathbb{R} \subset \mathbb{R}^2 \to \mathbb{R}$ como $h(x, y) = f(x) - y$. Pruebe:

    a) que h es derivable en el punto $(x_0, y_0) = (x_0, f(x_0)) \in I \times \mathbb{R}$

b) que $G_f = N_0(h)$ y que la recta tangente al conjunto de nivel $N_0(h)$ en el punto $(x_0, f(x_0))$, calculada de acuerdo a la definición 4.6, coincide con ser la recta tangente a la gráfica de f en el punto $(x_0, f(x_0))$, calculada de acuerdo a como lo hacía en su primer curso de cálculo.

11. Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ derivable en $\hat{x}_0 = (x_0, y_0) \in U$. Definimos $h : U \times \mathbb{R} \subset \mathbb{R}^3 \to \mathbb{R}$ como $h(x, y, z) = f(x, y) - z$. Pruebe:

    a) que h es derivable en el punto $(x_0, y_0, z_0) = (x_0, y_0, f(x_0, y_0)) \in U \times \mathbb{R}$

    b) que $G_f = N_0(h)$ y que el plano tangente al conjunto de nivel $N_0(h)$ en el punto (x_0, y_0, z_0), calculada de acuerdo a la definición 4.6, coincide con ser el plano tangente a la gráfica de f en el punto (x_0, y_0, z_0), calculado de acuerdo con la definición 4.4.

12. De un ejemplo de una función $f : \mathbb{R}^2 \to \mathbb{R}$ que sólo sea derivable en el $(0, 0)$.

13. Muestre que las derivadas parciales de la función definida en el ejemplo 4.7 no son continuas en el $(0, 0)$ (*sugerencia*: hay dos formas de hacer esto, una directa, mostrando que el límite de ambas derivadas en el $(0, 0)$ no existe, y una indirecta, usando los resultados probados en el texto).

14. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ tal que $\frac{\partial f}{\partial x_i}$ existe y es continua en cada punto $\hat{x} \in U$, para $i \in \{1, \ldots, n\}$. Si $x'_1, \ldots, x'_n$ es otro sistema de coordenadas (inducido por otra base ortonormal $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$), pruebe que $\frac{\partial f}{\partial x'_i}$ existe y es continua en cada punto $\hat{x} \in U$, para $i \in \{1, \ldots, n\}$.

15. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Pruebe que, si para toda $i \in \{1, \ldots, n\}$ se tiene que $\frac{\partial f}{\partial x_i}(\hat{x})$ existe para cada $\hat{x} \in B_r(\hat{x}_0)$ y $\frac{\partial f}{\partial x_i}$ está acotada en $B_r(\hat{x}_0)$, entonces f es continua en $\hat{x}_0$.

16. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$, $\gamma : (a,b) \subset \mathbb{R} \to \mathbb{R}^n$ tal que $\gamma(a,b) \subset U$, y $t_0 \in (a,b)$ tal que $\gamma(t_0) = \hat{x}_0$. Definimos $\varphi : (a,b) \subset \mathbb{R} \to \mathbb{R}$ como

$$\varphi(t) = \begin{cases} \frac{f(\gamma(t)) - f(\gamma(t_0)) - \nabla f(\gamma(t_0)) \cdot (\gamma(t) - \gamma(t_0))}{\|\gamma(t) - \gamma(t_0)\|} & \text{si } \gamma(t) - \gamma(t_0) \neq \hat{0} \\ 0 & \text{si } \gamma(t) - \gamma(t_0) = \hat{0} \end{cases}$$

Pruebe que, si f es derivable en $\hat{x}_0$ y γ es continua en t_0, entonces φ es continua en t_0.

17. Pruebe la proposición 4.14.

18. Pruebe el corolario 4.1.

19. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ continua. Use la regla de la cadena para demostrar que:

$$\frac{d}{dx}\left(\int_0^x f(x,y)dy\right) = f(x,x) + \int_0^x \frac{\partial f}{\partial x}(x,y)dy.$$

Sugerencia: defina

$$F(u,v) = \int_0^u f(v,y)dy$$

y dé por hecho que

$$\frac{\partial F}{\partial v}(u,v) = \int_0^u \frac{\partial f}{\partial x}(v,y)dy.$$

20. Sea $f : A \subset \mathbb{R}^n \to \mathbb{R}$. Se dice que f es homogénea de grado $p \in \mathbb{N}$ si $f(\lambda\hat{x}) = \lambda^p f(\hat{x})$ para toda $\lambda \in \mathbb{R}$ y toda $\hat{x} \in A$ tales que $\lambda\hat{x} \in A$. Pruebe que, si f es derivable en $\hat{x}_0 \in A$, entonces $\hat{x}_0 \cdot \nabla f(\hat{x}_0) = pf(\hat{x}_0)$ (este resultado es conocido como el teorema de Euler para funciones homogéneas).

21. Calcula el plano tangente a las superficies determinadas por las siguientes ecuaciones, en el punto indicado:

    a) $x^2 + 2y^2 + 3xz = 10$ en $(1, 2, 1/3)$

    b) $xe^z + yz = 1$ en $(1, 1, 0)$

    c) $z = \cos(x)\,\mathrm{sen}(y)$ en $(0, \pi/2, 1)$.

22. Sean $f, g : \mathbb{R}^3 \to \mathbb{R}$. Supóngase que f es derivable y que $\nabla f(\hat{x}) = g(\hat{x})\hat{x}$ para toda $\hat{x} \in \mathbb{R}^3$. Prueba que las esferas en $\mathbb{R}^3$ con centro en el origen están contenidas en los conjuntos de nivel de f, es decir, que f es constante sobre estas esferas.

23. Sea $f : U \subset \mathbb{R}^3 \to \mathbb{R}$ una función que está expresada en términos de las coordenadas cilíndricas (ρ, θ, z) de cada punto $\hat{x} \in U$. Si $\hat{x}_0 \in U$ tiene coordenadas cilíndricas (ρ_0, θ_0, z_0)

    a) defina $\frac{\partial f}{\partial \rho}$, $\frac{\partial f}{\partial \theta}$ y $\frac{\partial f}{\partial z}$ en el punto $\hat{x}_0$. Interprete el significado geométrico de estas derivadas

    b) relacione a cada una de las derivadas $\frac{\partial f}{\partial \rho}$, $\frac{\partial f}{\partial \theta}$ y $\frac{\partial f}{\partial z}$ en el punto $\hat{x}_0$ con alguna derivada direccional en el mismo punto

    c) para cada punto $\hat{x} \in U$ defina una base ortonormal de $\mathbb{R}^3$ en la cual se pueda expresar la derivada de f en $\hat{x}$ $(Df(\hat{x}))$ en términos de las derivadas definidas en el inciso anterior

    d) exprese a cada una de las derivadas $\frac{\partial f}{\partial \rho}$, $\frac{\partial f}{\partial \theta}$ y $\frac{\partial f}{\partial z}$ en el punto $\hat{x}_0$ en términos de las derivadas $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ y $\frac{\partial f}{\partial z}$ en el punto $\hat{x}_0$, y viceversa.

24. Repita el problema anterior, suponiendo ahora que f está expresada en términos de coordenadas esféricas (ρ, θ, φ).

25. Sean, $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^2 en U, y $\hat{u}, \hat{v} \in \mathbb{R}^n$ vectores unitarios. Pruebe que $D_{\hat{v}}(D_{\hat{u}} f)$ y $D_{\hat{u}}(D_{\hat{v}} f)$ existen y

$$D_{\hat{v}}(D_{\hat{u}} f)(\hat{x}) = D_{\hat{u}}(D_{\hat{v}} f)(\hat{x})$$

    para toda $\hat{x} \in U$.

26. *a*) Defina las derivadas parciales de orden dos en coordenadas polares (ver definición 4.8), en coordenadas cilíndricas (ver problema 23), y en coordenadas esféricas (ver problema 24).

    b) Pruebe que el teorema de las derivadas parciales cruzadas también se cumple para estas derivadas parciales de orden dos.

27. Sea

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{si } (x,y) \neq (0,0) \\ \\ 0 & \text{si } (x,y) = (0,0) \end{cases}$$

    a) calcule $\frac{\partial f}{\partial x}(x,y)$ y $\frac{\partial f}{\partial y}(x,y)$ para $(x,y) \neq (0,0)$

    b) pruebe que $\frac{\partial f}{\partial x}(0,0) = 0 = \frac{\partial f}{\partial y}(0,0)$

    c) pruebe que $\frac{\partial^2 f}{\partial y \partial x}(0,0) = -1$ y que $\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1$

    d) ¿este ejemplo contradice el resultado de la proposición 4.1? Justifique su respuesta.

28. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $i, j \in \{1, \ldots, n\}$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Pruebe que, si $\frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x})$ existe para toda $\hat{x} \in B_r(\hat{x}_0)$ y es continua en $\hat{x}_0$, y $\frac{\partial f}{\partial x_j}(\hat{x})$ existe para toda $\hat{x} \in B_r(\hat{x}_0)$, entonces $\frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0)$ existe y además

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x}_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\hat{x}_0).$$

29. Sean $F_1, \ldots, F_n : U \subset \mathbb{R}^n \to \mathbb{R}$ tal que F_i es de clase C^1 en U para $i = 1, \ldots, n$, y supóngase que existe $f : U \subset \mathbb{R}^n \to \mathbb{R}$ tal que $\frac{\partial f}{\partial x_i}(\hat{x}) = F_i(\hat{x})$ para toda $\hat{x} \in U$ $(i = 1, \ldots, n)$. Pruebe que $\frac{\partial F_i}{\partial x_j}(\hat{x}) = \frac{\partial F_j}{\partial x_i}(\hat{x})$ para toda $\hat{x} \in U$ $(i, j = 1, \ldots, n)$.

30. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^{k+1} en U. Pruebe que cualquier derivada parcial de orden k de f es (como función de $\mathbb{R}^n$ en $\mathbb{R}$) derivable en cualquier punto $\hat{x} \in U$.

31. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$. Pruebe que, si f es de clase C^{k+1} en U, entonces f es de clase C^k en U (para toda $k \in \mathbb{N}$).

32. Sea $f : \mathbb{R}^n \to \mathbb{R}$ de clase C^2 en $\mathbb{R}^n$ y $g : \mathbb{R} \to \mathbb{R}$ dos veces derivable. Pruebe que $\frac{\partial^2 (g \circ f)}{\partial x_j \partial x_i}(\hat{x})$ existe para toda $\hat{x} \in \mathbb{R}^n$ $(i, j = 1, \ldots, n)$ y dé una expresión para estas derivadas parciales en términos de f y g.

33. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $\hat{x}_0 \in U$ y $r > 0$ tales que f es de clase C^k en $B_r(\hat{x}_0) \subset U$. Si $\hat{u} = (u_1, \ldots, u_n)$ es tal que $0 < \|\hat{u}\| < r$, y $\gamma : (-r/\|\hat{u}\|, r/\|\hat{u}\|) \subset \mathbb{R} \to \mathbb{R}^n$ está dada por $\gamma(t) = \hat{x}_0 + t\hat{u}$, pruebe que la función

$$f \circ \gamma : (-r/\|\hat{u}\|, r/\|\hat{u}\|) \subset \mathbb{R} \to \mathbb{R}$$

es de clase C^k en $(-r/\|\hat{u}\|, r/\|\hat{u}\|)$ y que además

$$(f \circ \gamma)^{(m)}(t) = \sum_{i_1, \ldots, i_m = 1}^{n} \frac{\partial^m f}{\partial x_{i_m} \cdots \partial x_{i_1}} (\gamma(t))\, u_{i_1} \cdots u_{i_m}$$

para cada $t \in (-r/\|\hat{u}\|, r/\|\hat{u}\|)$ y cada $m \in \{1, \ldots, k\}$.

34. Sean $f, g : \mathbb{R} \to \mathbb{R}$ dos veces derivables.

    a) Definimos $h : \mathbb{R}^2 \to \mathbb{R}$ como $h(x, t) = f(x - at) + g(x + at)$ con $a \in \mathbb{R}$ una constante. Pruebe que:

$$a^2 \frac{\partial^2 h}{\partial x^2}(x, t) = \frac{\partial^2 h}{\partial t^2}(x, t) \text{ para toda } (x, t) \in \mathbb{R}^2.$$

    b) Definimos $h : \mathbb{R}^2 \to \mathbb{R}$ como $h(x, y) = xf(x + y) + yg(x + y)$. Pruebe que:

$$\frac{\partial^2 h}{\partial x^2}(x, y) + \frac{\partial^2 h}{\partial y^2}(x, y) = 2\frac{\partial^2 h}{\partial y \partial x}(x, y) \text{ para toda } (x, y) \in \mathbb{R}^2.$$

c) Definimos $h : \mathbb{R}^2 \to \mathbb{R}$ como $h(x, y) = f(x^2 + y^2)$. Pruebe que:

$$y^2 \frac{\partial^2 h}{\partial x^2}(x, y) - x^2 \frac{\partial^2 h}{\partial y^2}(x, y) + x \frac{\partial h}{\partial x}(x, y) - y \frac{\partial h}{\partial y}(x, y) = 0,$$

para toda $(x, y) \in \mathbb{R}^2$.

35. Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ y $\hat{x}_0 \in U$.

a) Pruebe que, si f es derivable en $\hat{x}_0$, entonces la función polinomial

$$P_2(x, y) = A(x - x_0)^2 + B(y - y_0)^2 + C(x - x_0)(y - y_0) + \frac{\partial f}{\partial x}(\hat{x}_0)(x - x_0) + \frac{\partial f}{\partial y}(\hat{x}_0)(y - y_0) + f(\hat{x}_0)$$

satisface que

$$\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - P_2(\hat{x})}{\|\hat{x} - \hat{x}_0\|} = 0$$

para cualesquiera A, B y C números reales.

b) Muestre con un ejemplo que el inciso anterior es falso si sólo suponemos que $\frac{\partial f}{\partial x}(\hat{x}_0)$ y $\frac{\partial f}{\partial y}(\hat{x}_0)$ existen.

36. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^N en U y $\hat{x}_0 \in U$. Definimos

$$P_N(x_1, \ldots, x_n) = f(\hat{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\hat{x}_0) x_i + \cdots + \frac{1}{N!} \sum_{i_1, \ldots, i_N = 1}^{n} \frac{\partial^N f}{\partial x_{i_1} \cdots \partial x_{i_N}}(\hat{x}_0) x_{i_1} \cdots x_{i_N}$$

Si $\hat{h} = (h_1, \ldots, h_n)$, pruebe que

$$\lim_{\hat{h} \to \hat{0}} \frac{f(\hat{x}_0 + \hat{h}) - P_N(\hat{h})}{\left\|\hat{h}\right\|^N} = 0$$

(*sugerencia*: use la expresión del residuo vista en la demostración del teorema de Taylor (para el caso $N - 1$)).

37. Use el polinomio de Taylor de grado dos de la función $f(x,y) = \cos(x+y)$ para calcular el siguiente límite:

$$\lim_{(x,y)\to(0,0)} \frac{1-\cos(x+y)}{(x^2+y^2)^{1/2}}.$$

Pruebe su respuesta.

38. Sea $f(x,y,z) = Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + G$. Pruebe que el residuo del polinomio de Taylor de orden dos de esta función siempre vale cero (en cualquier punto). Con base en lo anterior:

a) escriba el polinomio $x^2+y^2+z^2$ en potencias de $x-1, y-1$ y $z-3$

b) escriba el polinomio $x^2-2y^2+4z^2+xy-yz+xz+1$ en potencias de $x-1, y+1$ y z.

39. Sea $p(x_1,\ldots,x_n)$ una función polinomial de las variables $x_1,\ldots,x_n$. Decimos que p es homogénea de orden $l \in \mathbb{N}$, si

$$p(tx_1,\ldots,tx_n) = t^l p(x_1,\ldots,x_n)$$

para toda $t \in \mathbb{R}$ y todo $(x_1,\ldots,x_n) \in \mathbb{R}^n$. Pruebe que:

a) si $f : \mathbb{R} \to \mathbb{R}$ es tal que $f^{(k)}(0)$ existe y $f \circ p$ es de clase C^{kl} en $\mathbb{R}^n$, entonces

$$P_{kl,f\circ p,\hat{0}} = P_{k,f,0} \circ p$$

b) para $s \in \{1,\ldots,kl\}$, $s \neq il$, con $i \in \{1,\ldots,k\}$, se tiene que

$$\frac{\partial^s (f\circ p)}{\partial x_n^{i_n}\cdots\partial x_1^{i_1}}\left(\hat{0}\right) = 0$$

en donde $i_1,\ldots,i_n \in \mathbb{N}$ son tales que $i_1+\cdots+i_n = s$.

40. Usando el polinomio de Taylor de grado adecuado de la función e^t, y sin hacer demasiados cálculos, encuentre todas las derivadas parciales de orden tres, en el $(0,0)$, de la función $f(x,y) = (x^2+y^2)e^{x^2+y^2}$.

41. Por un procedimiento análogo al del problema anterior, encuentre el valor de todas las derivadas parciales de orden tres, en el $(0,0)$, de la función $f(x,y) = (e^{xy} - 1)\operatorname{sen}(x+y)$.

42. Sean $x_1, \ldots, x_n \in \mathbb{R}$ y $N \in \mathbb{N}$. Use el teorema de Taylor para probar que

$$(x_1 + \cdots + x_n)^N = \sum_{\substack{k_1+\cdots+k_n=N \\ 0 \leq k_i}} \begin{pmatrix} N \\ k_1 \cdots k_n \end{pmatrix} x_1^{k_1} \cdots x_n^{k_n}$$

donde

$$\begin{pmatrix} N \\ k_1 \cdots k_n \end{pmatrix} = \frac{N!}{k_1! \cdots k_n!}$$

(este número es conocido con el nombre de *coeficiente multinomial*).

43. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $H : V \subset \mathbb{R}^k \to \mathbb{R}^n$, U y V abiertos, $\hat{x}_0 \in U$ y $\hat{y}_0 \in V$ tales que $H(\hat{y}_0) = \hat{x}_0$. Pruebe que, si f alcanza un valor máximo (mínimo) local en $\hat{x}_0$ y H es continua en $\hat{y}_0$, entonces $f \circ H$ alcanza un valor máximo (mínimo) local en $\hat{y}_0$. ¿Este resultado sigue siendo cierto si H no es continua en $\hat{y}_0$? Pruebe su respuesta.

44. Sea $n \in \mathbb{N}$, $n \geq 2$, y $f(x,y) = ax^n + by^n$, con $ab \neq 0$. Muestre que el $(0,0)$ es el único punto crítico de f y determine su tipo en términos de a, b y n.

45. Sea $f(x,y,z) = ax^2 + by^2 + cz^2$, $abc \neq 0$. Muestre que el $(0,0,0)$ es el único punto crítico de f y determine su tipo en términos de a, b y c.

46. Pruebe el inciso 2 de la proposición 4.20.

47. Pruebe el inciso 1 de la proposición 4.21.

48. Sea f una función de clase C^2 en una vecindad de $\hat{x}_0 \in \mathbb{R}^n$. Pruebe que, si f alcanza un mínimo local en $\hat{x}_0$, entonces

$$\frac{\partial^2 f}{\partial x_i^2}(\hat{x}_0) \geq 0$$

para cada $i \in \{1, \ldots, n\}$ (*sugerencia*: use la proposición 4.21).

49. Sea $\hat{x}_0 = (x_1, \ldots, x_n)$ un punto crítico de una función f de clase C^2 en una vecindad de $\hat{x}_0 \in \mathbb{R}^n$, tal que

$$\frac{\partial^2 f}{\partial x_i^2}(\hat{x}_0)\frac{\partial^2 f}{\partial x_j^2}(\hat{x}_0) < 0$$

para algunas $i, j \in \{1, \ldots, n\}$, con $i \neq j$. Pruebe que $\hat{x}_0$ es un punto silla de f.

50. Sea $\hat{x}_0 = (x_0, y_0)$ un punto crítico de una función f de clase C^2 en una vecindad de $\hat{x}_0 \in \mathbb{R}^2$, tal que

$$\frac{\partial^2 f}{\partial x^2}(\hat{x}_0)\frac{\partial^2 f}{\partial y^2}(\hat{x}_0) - \left(\frac{\partial^2 f}{\partial y \partial x}(\hat{x}_0)\right)^2 < 0.$$

Pruebe que $\hat{x}_0$ es un punto silla de f (*sugerencia*: observe que la cantidad anterior está relacionada con el discriminante del polinomio de grado 2 que se obtiene en la variable m al evaluar $Hf_{\hat{x}_0}(1, m)$, para $m \in \mathbb{R}$).

51. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ definida como $f(x, y) = (y - 3x^2)(y - x^2)$. Pruebe que:

    a) el origen es un punto crítico de f

    b) si $a, b \in \mathbb{R}$ y $g(t) = (at, bt)$ con $t \in \mathbb{R}$, entonces $f \circ g$ tiene un mínimo local en $t = 0$

    c) el origen no es un mínimo local de f.

52. Sea $f(x, y) = -(x^2 - 1)^2 - (x^2 y - x - 1)^2$. Pruebe que:

    a) f sólo tiene dos puntos críticos

    b) ambos puntos críticos son máximos locales

    c) ¿se puede presentar una situación análoga para funciones de $\mathbb{R}$ en $\mathbb{R}$? Pruebe su respuesta.

53. Sea $f(x,y) = 3xe^y - x^3 - e^{3y}$. Pruebe que:

    a) f sólo tiene un punto crítico

    b) el punto crítico es un máximo local

    c) f no tiene un máximo global

    d) ¿se puede presentar una situación análoga para funciones de $\mathbb{R}$ en $\mathbb{R}$? Pruebe su respuesta.

54. Encuentre el máximo y el mínimo de la función $f(x,y) = xy - y + x - 1$ sobre el conjunto

$$A = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 2\}.$$

55. Sean $(x_1, y_1), \ldots, (x_n, y_n)$ puntos en $\mathbb{R}^2$ con $x_1 < x_2 < \ldots < x_n$. Pruebe que:

    a) la función

$$d(m,b) = \sum_{i=1}^{n} (y_i - (mx_i + b))^2$$

    alcanza un valor mínimo en $\mathbb{R}^2$. Encuentre los valores m_0 y b_0 para los cuales se alcanza este valor mínimo (la recta $y = m_0x + b_0$ es la que "mejor" aproxima a los puntos $(x_1, y_1), \ldots, (x_n, y_n)$ y al método que se usa para calcular m_0 y b_0 se le conoce como el **método de los cuadrados mínimos**).

    b) si m_0 y b_0 son los valores del inciso anterior, pruebe que

$$\sum_{i=1}^{n} (y_i - (m_0x_i + b_0)) = 0$$

    c) si $n = 2$ (es decir, sólo se toman dos puntos), entonces $y = m_0x + b_0$ es la recta que pasa por dichos puntos.

56. Encuentre los extremos de f relativos a S, donde:

a) $f(x,y) = x^2 - y^2$ y $S = \{(x, \cos(x)) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$

b) $f(x,y,z) = x^2 + y^2 + z^2$ y $S = \{(x,y,z) \in \mathbb{R}^3 \mid z \geq 2 + x^2 + y^2\}$.

57. Escriba el número 120 como suma de tres números, de modo que la suma de sus productos, tomados de dos en dos, sea máxima.

58. Una compañía planea fabricar cajas rectangulares cerradas con un volumen de 8 litros. El material para la base y la tapa cuesta el doble que el que se usa para los lados. Encuentre las dimensiones para las cuales el costo es mínimo.

59. Una ventana tiene la forma de un triángulo isósceles montado sobre un rectángulo. Si la base del rectángulo mide x, su altura mide y, los ángulos de la base del triángulo miden θ, y el perímetro de la ventana mide 4 metros, encuentre los valores de x, y y θ que hacen que el área de la ventana sea máxima.

60. Tres alelos (formas mutantes de genes) **A**, **B** y **O** determinan los cuatro tipos sanguíneos: **A**(**AA** o **AO**), **B** (**BB** o **BO**), **O** (**OO**) y **AB**. La ley de Hardy-Weinberg establece que la proporción P de individuos de una población que llevan dos alelos diferentes es $P = 2pq+2pr+2qr$, donde p, q y r son las proporciones de los alelos **A**, **B** y **O** que se presentan en dicha población, respectivamente. Use el hecho de que $p + q + r = 1$ para demostrar que $P \leq 2/3$.

61. Sea $P \in S = N_1(f) \subset \mathbb{R}^3$ con f de clase C^1 en $\mathbb{R}^3$. Supóngase que P es un punto donde se maximiza la distancia del origen a S. Pruebe que el vector que sale del origen y termina en P es perpendicular a S.

62. Sea A una matriz de 3×3, simétrica y diferente de la matriz cero. Definimos $f(x,y,z) = \frac{1}{2}(A(x,y,z)^t) \cdot (x,y,z)$ y $S = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$. Pruebe que:

    a) si $\hat{x}_0 \in S$ es un punto en donde f alcanza su valor máximo (o mínimo) sobre S, entonces $\hat{x}_0$ es un vector propio de A, es decir, que existe λ tal que $A\hat{x}_0 = \lambda\hat{x}_0$

b) existe $\hat{x}_0 \in S$ tal que $\hat{x}_0$ es un vector propio de A correspondiente a un valor propio distinto de cero, es decir, que existe $\lambda \neq 0$ tal que $A\hat{x}_0 = \lambda \hat{x}_0$.

63. Sean $a_1, \ldots, a_k$ números positivos. Use multiplicadores de Lagrange para probar que:

$$(a_1 \cdots a_k)^{1/k} \leq \frac{a_1 + \cdots + a_k}{k}.$$

64. Use multiplicadores de Lagrange para demostrar que

$$a_1 b_1 + \cdots + a_n b_n \leq (a_1^p + \cdots + a_n^p)^{1/p} \, (b_1^q + \cdots + b_n^q)^{1/q},$$

donde $a_i, b_i \geq 0$ para $i = 1, \ldots, n$, y $p, q > 1$ tales que $\frac{1}{p} + \frac{1}{q} = 1$ (esta desigualdad es conocida como *desigualdad de Hölder*).

65. Pruebe que

$$((a_1 + b_1)^p + \cdots + (a_n + b_n)^p)^{1/p} \leq (a_1^p + \cdots + a_n^p)^{1/p} + (b_1^p + \cdots + b_n^p)^{1/p},$$

donde $a_i, b_i \geq 0$ para $i = 1, \ldots, n$, y $p \geq 1$ (esta desigualdad es conocida como *desigualdad de Minkowski*).

Capítulo 5

La derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$

En este capítulo introduciremos el concepto de derivada para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$, que entre otras cosas, tendrá que coincidir con los conceptos que hemos desarrollado en los capítulos anteriores cuando $n = 1$ o $m = 1$. Además de definir el concepto de derivada para este tipo de funciones, formularemos y probaremos sus propiedades más importantes, algunas de las cuales generalizan a las correspondientes propiedades vistas en los capítulos 3 y 4. Por la razón anterior, mucho del material que veremos ahora estará basado en las ideas y conceptos de los capítulos anteriores (sobre todo del capítulo 4).

Para finalizar este capítulo (¡y este texto!) formularemos y probaremos dos teoremas, que sin duda son de los más importantes del Cálculo Diferencial de varias Variables: el Teorema de la Función Implícita y el Teorema de la Función Inversa.

5.1. La derivada

Como era de esperarse, la definición de derivada de una función $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ en un punto $\hat{x}_0 \in U$ está basada en la misma idea que se usó para

las funciones de $\mathbb{R}^n$ en $\mathbb{R}$: la existencia de "la mejor aproximación lineal a f alrededor del punto $\hat{x}_0$".

Para empezar notemos que, como en el capítulo anterior, si $L : \mathbb{R}^n \to \mathbb{R}^m$ es una función lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$, entonces $L(\hat{x}-\hat{x}_0)+f(\hat{x}_0)$ es una función afín de $\mathbb{R}^n$ en $\mathbb{R}^m$ cuyo valor en $\hat{x}_0$ coincide con el valor de f en $\hat{x}_0$. Ahora, como el lector estará de acuerdo, decir que una función afín de éstas "se parece mucho a f alrededor del punto $\hat{x}_0$", significará que la diferencia de estas funciones, evaluada en puntos $\hat{x}$ "cercanos" a $\hat{x}_0$, tiende a $\hat{0}$ más rápido de lo que la diferencia $\hat{x} - \hat{x}_0$ también tiende a $\hat{0}$.

De manera más precisa, la discusión anterior queda plasmada en la siguiente

Definición 5.1 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ y $\hat{x}_0 \in U$. Decimos que f es derivable en $\hat{x}_0$ si existe $L : \mathbb{R}^n \to \mathbb{R}^m$ una función lineal tal que*

$$\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - (L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} = \hat{0},$$

o lo que es equivalente (problema 33 del capítulo 2), si

$$\lim_{\hat{x}\to\hat{x}_0} \frac{\|f(\hat{x}) - (L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))\|}{\|\hat{x} - \hat{x}_0\|} = 0.$$

Lo que ahora procede es dar un ejemplo de cómo calcular la derivada que acabamos de definir, pero que ésta no sea de una función del tipo de las que ya vimos en el capítulo 3 o en el capítulo 4, es decir, la derivada de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ en donde $n > 1$ y $m > 1$.

Observación 5.1 *Antes de ello, y en beneficio del propio ejemplo, conviene hacer notar que la definición anterior es en efecto una generalización de las definiciones de derivada dadas en los capítulos antes mencionados.*

Que la definición 4.3 del capítulo 4 es un caso particular de la definición 5.1, es un hecho que el lector puede verificar con sólo leer su enunciado. Para verificar la equivalencia entre la definición 5.1 y la definición 3.2 dada en el capítulo 3, el lector tendrá que recurrir al problema 9 de este capítulo, del cual se desprende fácilmente lo que aquí afirmamos.

Una vez dicho lo anterior, procedemos a dar el siguiente

Ejemplo 5.1 *Sea* $f : \mathbb{R}^2 \to \mathbb{R}^3$ *definida como*

$$f(x, y) = (x^2 \cos(y), x^2 \operatorname{sen}(y), x^2).$$

Dado que la definición 5.1, como ya mencionamos, es una generalización de los conceptos de derivada que vimos en los capítulos 3 y 4, por la proposición 3.1 del primero de estos dos capítulos, debemos sospechar que la derivabilidad (¡y la derivada!) de f *está determinada por la derivabilidad (¡y la derivada!) de cada una de sus funciones coordenadas, funciones que son del tipo de las que vimos en el capítulo 4.*

En virtud de lo anterior, todo parece indicar que para obtener la derivada de f *en un punto* $(x_0, y_0) \in \mathbb{R}^2$*, será necesario calcular la derivada de sus funciones coordenadas* $f_1(x, y) = x^2 \cos(y)$*,* $f_2(x, y) = x^2 \operatorname{sen}(y)$ *y* $f_3(x, y) = x^2$ *en este punto.*

Ahora, por los resultados obtenidos en el capítulo 4, sabemos que las funciones lineales L_1*,*L_2 *y* L_3 *de* $\mathbb{R}^2$ *en* $\mathbb{R}$ *asociadas a las matrices de* 1×2

$$\begin{bmatrix} \frac{\partial f_1}{\partial x}(x_0, y_0) & \frac{\partial f_1}{\partial y}(x_0, y_0) \end{bmatrix} = \begin{bmatrix} 2x_0 \cos(y_0) & -x_0^2 \operatorname{sen}(y_0) \end{bmatrix},$$
$$\begin{bmatrix} \frac{\partial f_2}{\partial x}(x_0, y_0) & \frac{\partial f_2}{\partial y}(x_0, y_0) \end{bmatrix} = \begin{bmatrix} 2x_0 \operatorname{sen}(y_0) & x_0^2 \cos(y_0) \end{bmatrix}$$

y

$$\begin{bmatrix} \frac{\partial f_3}{\partial x}(x_0, y_0) & \frac{\partial f_3}{\partial y}(x_0, y_0) \end{bmatrix} = \begin{bmatrix} 2x_0 & 0 \end{bmatrix}$$

son las funciones lineales que más se le "parecen" a las respectivas funciones coordenadas alrededor del punto (x_0, y_0)*. Por esta razón, la función lineal* L *de* $\mathbb{R}^2$ *en* $\mathbb{R}^3$ *que tiene como funciones coordenadas a las funciones* L_1*,*L_2 *y* L_3*, y cuya matriz asociada (de* 3×2*) está dada por*

$$\begin{bmatrix} \frac{\partial f_1}{\partial x}(x_0, y_0) & \frac{\partial f_1}{\partial y}(x_0, y_0) \\ \frac{\partial f_2}{\partial x}(x_0, y_0) & \frac{\partial f_2}{\partial y}(x_0, y_0) \\ \frac{\partial f_3}{\partial x}(x_0, y_0) & \frac{\partial f_3}{\partial y}(x_0, y_0) \end{bmatrix} = \begin{bmatrix} 2x_0 \cos(y_0) & -x_0^2 \operatorname{sen}(y_0) \\ 2x_0 \operatorname{sen}(y_0) & x_0^2 \cos(y_0) \\ 2x_0 & 0 \end{bmatrix}$$

sin duda es la mejor candidata para satisfacer la condición de la definición 5.1.

En efecto, de la definición 4.3 del capítulo 4 sabemos que para cada $i \in \{1, 2, 3\}$ *se satisface que*

$$\lim_{(x,y)\to(x_0,y_0)} \frac{f_i(x,y) - (L_i(x - x_0, y - y_0) + f_i(x_0, y_0))}{\|(x - x_0, y - y_0)\|} = 0.$$

Por la proposición 2.7 del capítulo 2 podemos concluir que

$$\lim_{(x,y)\to(x_0,y_0)} \frac{f(x,y) - (L(x - x_0, y - y_0) + f(x_0, y_0))}{\|(x - x_0, y - y_0)\|} = \hat{0},$$

de modo que f *es derivable en el punto* (x_0, y_0).

Tomaremos el ejemplo anterior como punto de partida para obtener algunas conclusiones sobre las funciones de $\mathbb{R}^2$ en $\mathbb{R}^3$. Empezaremos por aquellas que son de carácter geométrico.

5.1.1. Elementos básicos acerca de superficies

Como seguramente el lector recordará, en el capítulo 2 mencionamos que desde un punto de vista geométrico, lo importante de las funciones de $\mathbb{R}^2$ en $\mathbb{R}^3$ es su imagen. Así mismo, recordará que el concepto de derivada para funciones de $\mathbb{R}$ en $\mathbb{R}^3$ nos fue útil para definir la recta tangente, en un punto, de un conjunto que se puede ver como la imagen de una función de este tipo.

Ahora veremos que si un conjunto $S \subset \mathbb{R}^3$ se puede obtener como la imagen de una función, que en esta subsección denotaremos como $\sigma : U \subset \mathbb{R}^2 \to \mathbb{R}^3$, la cual es derivable en un punto $\hat{x}_0 \in U$, y esta derivada cumple con otra condición que precisaremos más adelante, entonces podremos definir lo que llamaremos el plano tangente a S en el punto $\sigma(\hat{x}_0)$.

Sea pues $S \subset \mathbb{R}^3$ un conjunto que se puede ver como la imagen de una función

$$\sigma = (\sigma_1, \sigma_2, \sigma_3) : U \subset \mathbb{R}^2 \to \mathbb{R}^3,$$

la cual es derivable en un punto $\hat{x}_0 = (x_0, y_0) \in U$. Dado que estamos suponiendo que U es un conjunto abierto, entonces existe $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. De esta forma, podemos definir las funciones

$$\gamma_x, \gamma_y : (-r, r) \subset \mathbb{R} \to \mathbb{R}^3$$

como

$$\begin{aligned}\gamma_x(t) &= \sigma(x_0 + t, y_0) \\ &= (\sigma_1(x_0 + t, y_0), \sigma_2(x_0 + t, y_0), \sigma_3(x_0 + t, y_0))\end{aligned}$$

y

$$\begin{aligned}\gamma_y(t) &= \sigma(x_0, y_0 + t) \\ &= (\sigma_1(x_0, y_0 + t), \sigma_2(x_0, y_0 + t), \sigma_3(x_0, y_0 + t)) .\end{aligned}$$

Afirmamos que estas funciones son derivables en $t = 0$.

En efecto, de acuerdo con la definición 3.2 del capítulo 3, se tiene que

$$\begin{aligned}\gamma_x'(0) &= \lim_{t\to 0} \frac{\gamma_x(t) - \gamma_x(0)}{t - 0} \\ &= \lim_{t\to 0} \left(\frac{\sigma_1(x_0 + t, y_0) - \sigma_1(x_0, y_0)}{t}, \frac{\sigma_2(x_0 + t, y_0) - \sigma_2(x_0, y_0)}{t}, \right. \\ &\qquad \left. \frac{\sigma_3(x_0 + t, y_0) - \sigma_3(x_0, y_0)}{t} \right) \\ &= \left(\frac{\partial \sigma_1}{\partial x}(x_0, y_0), \frac{\partial \sigma_2}{\partial x}(x_0, y_0), \frac{\partial \sigma_3}{\partial x}(x_0, y_0) \right)\end{aligned}$$

y

$$\begin{aligned}\gamma_y'(0) &= \lim_{t\to 0} \frac{\gamma_y(t) - \gamma_y(0)}{t - 0} \\ &= \lim_{t\to 0} \left(\frac{\sigma_1(x_0, y_0 + t) - \sigma_1(x_0, y_0)}{t}, \frac{\sigma_2(x_0, y_0 + t) - \sigma_2(x_0, y_0)}{t}, \right. \\ &\qquad \left. \frac{\sigma_3(x_0, y_0 + t) - \sigma_3(x_0, y_0)}{t} \right) \\ &= \left(\frac{\partial \sigma_1}{\partial y}(x_0, y_0), \frac{\partial \sigma_2}{\partial y}(x_0, y_0), \frac{\partial \sigma_3}{\partial y}(x_0, y_0) \right) .\end{aligned}$$

Por lo anterior, si se satisface que

$$\gamma_x'(0) = \left(\frac{\partial \sigma_1}{\partial x}(x_0, y_0), \frac{\partial \sigma_2}{\partial x}(x_0, y_0), \frac{\partial \sigma_3}{\partial x}(x_0, y_0) \right) \neq \hat{0}$$

y

$$\gamma_y'(0) = \left(\frac{\partial \sigma_1}{\partial y}(x_0, y_0), \frac{\partial \sigma_2}{\partial y}(x_0, y_0), \frac{\partial \sigma_3}{\partial y}(x_0, y_0) \right) \neq \hat{0},$$

entonces $\gamma_x((-r, r))$ y $\gamma_y((-r, r))$ son dos curvas contenidas en S que tienen recta tangente en el punto $\sigma(\hat{x}_0)$.

Ahora, si además los vectores $\gamma_x'(0)$ y $\gamma_y'(0)$ son linealmente independientes, el conjunto dado por

$$\left\{ \sigma(\hat{x}_0) + t\gamma_x'(0) + s\gamma_y'(0) \in \mathbb{R}^3 \mid t, s \in \mathbb{R} \right\}$$

representa un plano que tiene todo el aspecto de ser tangente a S en $\sigma(\hat{x}_0)$ (ver figura 5.1).

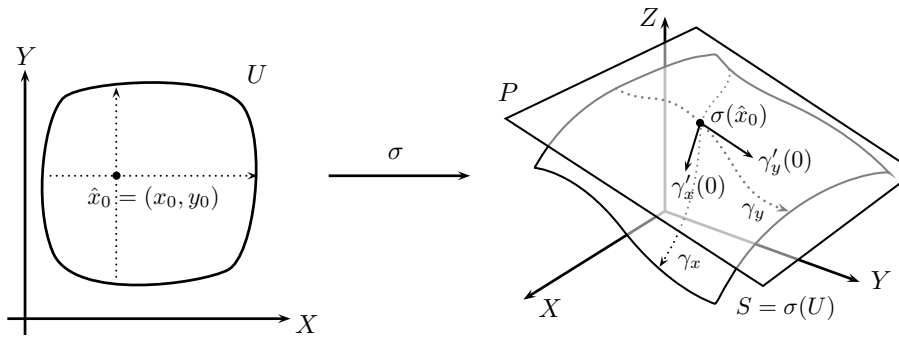


Figura 5.1: Si los vectores $\gamma_x'(0)$ y $\gamma_y'(0)$ son linealmente independientes, entonces el plano P generado por ellos, trasladado al punto $\sigma(\hat{x}_0)$, es tangente a la superficie $S = \sigma(U)$.

Con base en la discusión anterior, definiremos los siguientes conceptos.

Definición 5.2 *Sea $S \subset \mathbb{R}^3$.*

1. *Decimos que S es una superficie, si existen $\sigma = (\sigma_1, \sigma_2, \sigma_3) : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ derivable en cada punto de U, y $A \subset U$ tales que $\sigma(A) = S$. En este caso decimos que σ es* una *parametrización de S.*

2. *Si $\hat{x}_0 \in A$ y los vectores*

$$\frac{\partial \sigma}{\partial x}(\hat{x}_0) := \left(\frac{\partial \sigma_1}{\partial x}(\hat{x}_0), \frac{\partial \sigma_2}{\partial x}(\hat{x}_0), \frac{\partial \sigma_3}{\partial x}(\hat{x}_0) \right)$$

y

$$\frac{\partial \sigma}{\partial y}(\hat{x}_0) := \left(\frac{\partial \sigma_1}{\partial y}(\hat{x}_0), \frac{\partial \sigma_2}{\partial y}(\hat{x}_0), \frac{\partial \sigma_3}{\partial y}(\hat{x}_0) \right)$$

son linealmente independientes, decimos que el plano $P \subset \mathbb{R}^3$ definido paramétricamente como

$$P := \left\{ \sigma(\hat{x}_0) + t\frac{\partial \sigma}{\partial x}(\hat{x}_0) + s\frac{\partial \sigma}{\partial y}(\hat{x}_0) \in \mathbb{R}^3 \mid t, s \in \mathbb{R} \right\}$$

es el plano tangente a S en el punto $\sigma(\hat{x}_0)$, y que S es suave en $\sigma(\hat{x}_0)$.

Con relación a esta nueva definición de plano tangente en un punto de una superficie, es importante hacer algunas observaciones. La primera de ellas es que, aún cuando una superficie $S \subset \mathbb{R}^3$ es la imagen de alguna función derivable, esto no significa que el conjunto S no pueda tener "picos". Una prueba de este hecho nos lo proporciona la función

$$\sigma(x, y) = (x^2 \cos(y), x^2 \operatorname{sen}(y), x^2)$$

del ejemplo 5.1 la cual, ya sabemos, es derivable en cualquier punto $(x, y) \in \mathbb{R}^2$.

Como el lector podrá verificar fácilmente, el conjunto $S = \sigma(\mathbb{R}^2)$ es la parte superior del cono determinado por la ecuación cartesiana $u^2 + v^2 = w^2$ (figura 5.2), el cual claramente tiene un "pico" en el punto $\sigma(0, 0) = (0, 0, 0)$.

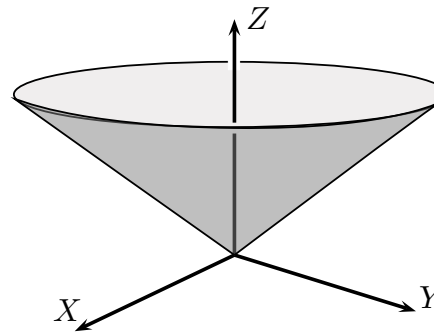


Figura 5.2: El cono determinado por la ecuación cartesiana $u^2 + v^2 = w^2$ tiene un "pico" en el punto $(0, 0, 0)$.

Una segunda observación es que no cualquier parametrización de una superficie $S \subset \mathbb{R}^3$ es útil para calcular su plano tangente en algún punto (del mismo modo que sucedió en el caso de las curvas).

El conjunto $S \subset \mathbb{R}^3$ y el par de parametrizaciones de éste que daremos en el siguiente ejemplo nos proporcionan una prueba de este hecho.

Ejemplo 5.2 *Sea $S \subset \mathbb{R}^3$ el siguiente conjunto:*

$$S = \left\{(u, v, w) \in \mathbb{R}^3 \mid w = u^2 + v^2\right\}.$$

La figura 5.3 muestra un esbozo de S, el cual corresponde a un paraboloide. A continuación daremos un par de parametrizaciones de este conjunto.

1. *Sea $\sigma : \mathbb{R}^2 \to \mathbb{R}^3$ definida como $\sigma(x, y) = \left(x, y, x^2 + y^2\right)$. Como el lector podrá verificar fácilmente, se tiene que $S = \sigma(\mathbb{R}^2)$. Por otra parte, procediendo como en el ejemplo 5.1, se concluye que σ es derivable para cualquier $(x, y) \in \mathbb{R}^2$ y además que*

$$\frac{\partial \sigma}{\partial x}(x, y) = (1, 0, 2x)$$

y

$$\frac{\partial \sigma}{\partial y}(x, y) = (0, 1, 2y).$$

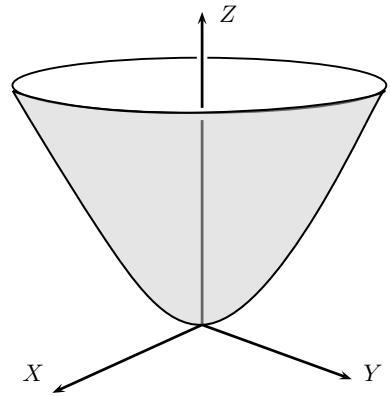


Figura 5.3: El paraboloide determinado por la ecuación cartesiana $u^2+v^2=w$ tiene como plano tangente en el punto $(0,0,0)$ al plano XY.

Por lo tanto, para el punto $(0,0)$ *se tiene que*

$$\frac{\partial \sigma}{\partial x}(0,0) = (1,0,0)$$

y

$$\frac{\partial \sigma}{\partial y}(0,0) = (0,1,0)\,,$$

los cuales son vectores linealmente independientes. De esta forma, el conjunto

$$\begin{aligned} P &= \left\{ \sigma(0,0) + t\frac{\partial \sigma}{\partial x}(0,0) + s\frac{\partial \sigma}{\partial y}(0,0) \in \mathbb{R}^3 \mid t,s \in \mathbb{R} \right\} \\ &= \left\{ (0,0,0) + t(1,0,0) + s(0,1,0) \in \mathbb{R}^3 \mid t,s \in \mathbb{R} \right\} \\ &= \left\{ (t,s,0) \in \mathbb{R}^3 \mid t,s \in \mathbb{R} \right\} \end{aligned}$$

sí resulta ser un plano (el plano XY*), que sin duda es el plano tangente a* S *en el punto* $\sigma(0,0) = (0,0,0)$.

2. *Sea ahora* $\sigma : \mathbb{R}^2 \to \mathbb{R}^3$ *definida como* $\sigma(x,y) = \big(x\cos(y), x\,\mathrm{sen}(y), x^2\big)$. *Nuevamente el lector podrá verificar fácilmente que* $S = \sigma(\mathbb{R}^2)$. *Por*

otra parte, para esta función también es sencillo mostrar que es derivable para cualquier $(x, y) \in \mathbb{R}^2$ *y además, que*

$$\frac{\partial \sigma}{\partial x}(x, y) = (\cos(y), \operatorname{sen}(y), 2x)$$

y

$$\frac{\partial \sigma}{\partial y}(x, y) = (-x \operatorname{sen}(y), x \cos(y), 0)\,.$$

Por lo tanto, para el punto $(0, 0)$ *se tiene que*

$$\frac{\partial \sigma}{\partial x}(0, 0) = (1, 0, 0)$$

y

$$\frac{\partial \sigma}{\partial y}(0, 0) = (0, 0, 0)\,,$$

los cuales no son vectores linealmente independientes, de tal forma que el conjunto

$$\begin{aligned} P &= \left\{ \sigma(0, 0) + t\frac{\partial \sigma}{\partial x}(0, 0) + s\frac{\partial \sigma}{\partial y}(0, 0) \in \mathbb{R}^3 \mid t, s \in \mathbb{R} \right\} \\ &= \left\{ (0, 0, 0) + t(1, 0, 0) + s(0, 0, 0) \in \mathbb{R}^3 \mid t, s \in \mathbb{R} \right\} \\ &= \left\{ (t, 0, 0) \in \mathbb{R}^3 \mid t, s \in \mathbb{R} \right\} \end{aligned}$$

no resulta ser un plano.

Concluimos nuestras observaciones acerca de la definición de plano tangente a una superficie recordando que en el capítulo 4 se definió este mismo concepto para dos objetos geométricos diferentes; la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$ (definición 4.4), y el conjunto de nivel de una función de $\mathbb{R}^3$ en $\mathbb{R}$ (definición 4.7). En el problema 11 de este mismo capítulo, el lector probó que la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$ siempre se puede obtener como el conjunto de nivel de una cierta función de $\mathbb{R}^3$ en $\mathbb{R}$, y que el plano tangente que se obtiene usando cualquiera de las definiciones antes mencionadas es el mismo. Algo análogo probará el lector en el problema 1 de

este capítulo; la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$ siempre se puede obtener como la imagen de una cierta función $\mathbb{R}^2$ en $\mathbb{R}^3$, y el plano tangente que se obtiene usando la definición 5.2 es el mismo que se obtiene usando la definición 4.4.

Como mencionamos en el capítulo 4, en este capítulo probaremos el Teorema de la Función Implícita, del cual podremos deducir (bajo ciertas hipótesis) que todo conjunto de nivel de una función de $\mathbb{R}^3$ en $\mathbb{R}$ se puede obtener (al menos "localmente") como la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$ y, por lo dicho en el párrafo anterior, entonces también como la imagen de una cierta función de $\mathbb{R}^2$ en $\mathbb{R}^3$.

En el problema 21 de este capítulo el lector probará la afirmación anterior, y además que el plano tangente calculado de acuerdo con la definición 5.2 es el mismo que se obtiene usando la definición 4.7.

5.2. Propiedades de la derivada

Seguramente el lector estará de acuerdo en que, para determinar la derivabilidad de la función del ejemplo 5.1, fue muy importante saber que sus funciones coordenadas eran derivables. Esto sin duda nos lleva a concluir que en general, para deducir la derivabilidad de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ en un punto, es suficiente con que sus funciones coordenadas lo sean en ese mismo punto.

La buena noticia es que la afirmación recíproca también es cierta, es decir, si una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ es derivable en un punto, sus funciones coordenadas también deben ser derivables en ese punto. Este hecho es el primer criterio importante que veremos para determinar la derivabilidad (¡o no derivabilidad!) de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ y lo dejamos plasmado en la siguiente

Proposición 5.1 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $\hat{x}_0 \in U$ y $f_1, \ldots, f_m : U \subset \mathbb{R}^n \to \mathbb{R}$ funciones coordenadas de f (en una base ortonormal $\{\hat{e}_1, \ldots, \hat{e}_m\}$ de $\mathbb{R}^m$). La función f es derivable en $\hat{x}_0$ si y sólo si f_i es derivable en $\hat{x}_0$ para cada $i \in \{1, \ldots, m\}$.*

Demostración. Sea $L : \mathbb{R}^n \to \mathbb{R}^m$ una función lineal y supongamos que $L_i : \mathbb{R}^n \to \mathbb{R}$ son las funciones coordenadas de L (que por lo tanto también son lineales) determinadas por la misma base ortonormal de $\mathbb{R}^m$ que determina a las funciones coordenadas de f.

Dado que la $i-$ésima coordenada del vector

$$\frac{1}{\|\hat{x} - \hat{x}_0\|} \left(f(\hat{x}) - (L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))\right) = \frac{f(\hat{x}) - (L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|}$$

está dada por

$$\frac{f_i(\hat{x}) - (L_i(\hat{x} - \hat{x}_0) + f_i(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|}$$

para cada $i \in \{1, \ldots, m\}$, por la proposición 2.7 (del capítulo 2), se sigue que

$$\lim_{\hat{x} \to \hat{x}_0} \frac{f(\hat{x}) - (L(\hat{x} - \hat{x}_0) + f(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} = \hat{0}$$

si y sólo si

$$\lim_{\hat{x} \to \hat{x}_0} \frac{f_i(\hat{x}) - (L_i(\hat{x} - \hat{x}_0) + f_i(\hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} = 0,$$

de donde las afirmaciones de esta proposición se concluyen inmediatamente. ■

Muchos resultados y propiedades de la derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ serán una consecuencia inmediata de la proposición anterior (y de los correspondientes resultados y propiedades para funciones de $\mathbb{R}^n$ en $\mathbb{R}$), y para empezar, la usaremos para probar la unicidad de la derivada de las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$. En efecto, como por la proposición 4.9 del capítulo 4 sabemos que la derivada de una función de $\mathbb{R}^n$ en $\mathbb{R}$ es única, por la proposición anterior concluimos que esto mismo sucede para las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$. Por esta razón, de aquí en adelante hablaremos de *la* derivada de f en $\hat{x}_0$ y la denotamos por $Df(\hat{x}_0)$ (igual que en el caso de las funciones de $\mathbb{R}^n$ en $\mathbb{R}$); es decir, $Df(\hat{x}_0)$ designará a la función lineal que al trasladarla al punto $\hat{x}_0$, construyendo la función $Df(\hat{x}_0)(\hat{x} - \hat{x}_0) + f(\hat{x}_0)$, obtenemos la función lineal afín que más se le parece a f alrededor del punto $\hat{x}_0$. Por otra

parte, como seguramente el lector recordará de su curso de Álgebra Lineal, toda función lineal $L : \mathbb{R}^n \to \mathbb{R}^m$ se puede representar por una matriz de $m \times n$ con entradas reales (que depende de las bases que se elijan para $\mathbb{R}^n$ y $\mathbb{R}^m$), y que se construye de la siguiente manera: Si $L : \mathbb{R}^n \to \mathbb{R}^m$ es una función lineal, $\{\hat{e}_1, \ldots, \hat{e}_n\}$ y $\{\tilde{e}_1, \ldots, \tilde{e}_m\}$ son bases ortonormales de $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente, y $\hat{x} \in \mathbb{R}^n$ es tal que

$$\begin{aligned} \hat{x} &= (x_1, \ldots, x_n) \\ &= x_1\hat{e}_1 + \cdots + x_n\hat{e}_n, \end{aligned}$$

entonces

$$\begin{aligned} L(\hat{x}) &= L\left(x_1\hat{e}_1 + \cdots + x_n\hat{e}_n\right) \\ &= x_1 L\left(\hat{e}_1\right) + \cdots + x_n L\left(\hat{e}_n\right) \end{aligned}$$

De esta forma, que si

$$L\left(\hat{e}_i\right) = a_{1i}\tilde{e}_1 + \cdots + a_{mi}\tilde{e}_m$$

para cada $i \in \{1, \ldots, n\}$, se tiene que

$$\begin{aligned} L(\hat{x}) &= x_1 L\left(\hat{e}_1\right) + \cdots + x_n L\left(\hat{e}_n\right) \\ &= x_1\left(a_{11}\tilde{e}_1 + \cdots + a_{m1}\tilde{e}_m\right) + \\ &\vdots \\ &\quad + x_n\left(a_{1n}\tilde{e}_1 + \cdots + a_{mn}\tilde{e}_m\right) \\ &= (x_1, \ldots, x_n) \cdot (a_{11}, \ldots, a_{1n})\,\tilde{e}_1 + \\ &\vdots \\ &\quad + (x_1, \ldots, x_n) \cdot (a_{m1}, \ldots, a_{mn})\,\tilde{e}_m \end{aligned}$$

de donde, usando matrices, obtenemos que

$$L(\hat{x}) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Por tanto, se tiene que la matriz

$$\beta = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

representa a la función L (en las bases $\{\hat{e}_1, \ldots, \hat{e}_n\}$ y $\{\tilde{e}_1, \ldots, \tilde{e}_m\}$).

Obsérvese que las matrices β_i de $1 \times n$ dadas por

$$\beta_i = \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix}$$

para cada $i \in \{1, \ldots, m\}$ representarán a sus correspondientes funciones coordenadas L_i, y recíprocamente, si la matriz β_i representa a la función lineal L_i, para cada $i \in \{1, \ldots, m\}$, entonces la matriz β representa a la función lineal L. De lo anterior, y de la identidad 4.17 del capítulo 4 tendremos que, si $f_1, \ldots, f_m$ son las funciones coordenadas de la función f (las cuales, reiteramos que dependen de la base de $\mathbb{R}^m$ que se elija), entonces la derivada de f en $\hat{x}_0$ ($Df(\hat{x}_0)$) estará representada por la matriz

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\hat{x}_0) \end{bmatrix}, \tag{5.1}$$

a la que se le conococe con el nombre de *matriz jacobiana* [1], y no sin cierto abuso de notación (¡nuevamente!) escribiremos que

$$Df(\hat{x}_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\hat{x}_0) \end{bmatrix}, \tag{5.2}$$

[1] Llamada así en honor de Carl Gustav Jacob Jacobi (10 de diciembre de 1804 en Potsdam, Prusia, actual Alemania -18 de febrero de 1851 en Berlín). Autor muy prolífico, contribuyó en varios campos de la matemática, principalmente en el área de las funciones elípticas, el álgebra, la teoría de números y las ecuaciones diferenciales. También destacó en su labor pedagógica, por la que se le ha considerado el profesor más estimulante de su tiempo. (fuente: Wikipedia).

Sin duda una condición que es necesaria y suficiente resulta ser muy útil (como la que se da en la proposición 5.1), sin embargo, las que sólo son necesarias o sólo son suficientes, también lo son. Y aprovechando que conocemos dos de este tipo de propiedades para funciones de $\mathbb{R}^n$ en $\mathbb{R}$, enunciaremos sus equivalentes para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$.

La primera de ellas es una condición (o consecuencia) necesaria de la derivabilidad de una función en un punto, y es una consecuencia inmediata de las proposiciones 5.1, 4.12 y 2.11.

Proposición 5.2 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$. Si f es derivable en $\hat{x}_0 \in U$, entonces f es continua en $\hat{x}_0$.*

La otra propiedad que es muy importante mencionar, y que es una condición suficiente para la derivabilidad de una función en un punto, es una consecuencia inmediata de las proposiciones 5.1 y 4.13.

Proposición 5.3 *Sean, $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $\hat{x}_0 \in U$ y $r > 0$ tal que $B_r(\hat{x}_0) \subset U$. Si $f_1, \ldots, f_m$ son funciones coordenadas de f tales que $\frac{\partial f_j}{\partial x_i}(\hat{x})$ existe para cada $\hat{x} \in B_r(\hat{x}_0)$ y $\frac{\partial f_j}{\partial x_i}$ es continua en $\hat{x}_0$, para cada $i \in \{1, \ldots, n\}$ y para cada $j \in \{1, \ldots, m\}$, entonces f es derivable en $\hat{x}_0$.*

Como vimos en el capítulo 4, la hipótesis de la proposición 4.13 (análoga a la anterior) dio lugar al concepto de función de clase C^k, concepto que jugó un papel muy importante en el tema de aproximación polinomial, y que ahora vamos a generalizar, apoyados justo en la proposición 5.1, a las funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ en la siguiente

Definición 5.3 *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $f_1, \ldots, f_m$ funciones coordenadas de f, y $k \in \mathbb{N}$. Decimos que f es una función de clase C^k en U si cada f_j es una función de clase C^k en U, para $j \in \{1, \ldots, m\}$. Es decir, si existen todas las derivadas parciales de orden k de f_j en cada punto de U, y además estas derivadas parciales son continuas en cada punto de U, para $j \in \{1, \ldots, m\}$.*

Con respecto a la definición anterior, hay que mencionar lo siguiente: por el problema 14 del capítulo 4 y la identidad 5.3 que probaremos en

la siguiente subsección, el hecho de que una función sea de clase C^k en un conjunto (abierto) es independiente tanto del sistema coordenado que estemos usando para representar a cada punto $\hat{x} \in \mathbb{R}^n$, como del que estemos usando para representar a cada punto $\hat{y} \in \mathbb{R}^m$. Es decir, este concepto es independiente de las variables coordenadas $x_1, \ldots, x_n$ y de las funciones coordenadas $f_1, \ldots, f_m$.

Como mencionamos anteriormente, con base en el concepto de función de clase C^k (para $k = 1$) podemos establecer el siguiente resultado, cuya prueba es una consecuencia inmediata de la proposición 4.17 del capítulo 4 y de la proposición 5.1.

Proposición 5.4 *Si $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ es de clase C^1 en U, entonces f es derivable para toda $\hat{x} \in U$.*

Las propiedades de la derivada de funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$ relacionadas con la aritmética de éstas también se obtienen de manera inmediata a partir de la proposición 5.1 y de las correspondientes propiedades para funciones de $\mathbb{R}^n$ en $\mathbb{R}$. Por esta razón, no las probaremos y sólo las dejaremos formuladas en la siguiente

Proposición 5.5 *Sean $f, g : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $\hat{x}_0 \in U$ y $\alpha \in \mathbb{R}$. Si f y g son derivables en $\hat{x}_0$, entonces:*

1. *$f + g$ es derivable en $\hat{x}_0$ y además*

$$D(f + g)(\hat{x}_0) = Df(\hat{x}_0) + Dg(\hat{x}_0)$$

2. *αf es derivable en $\hat{x}_0$ y además*

$$D(\alpha f)(\hat{x}_0) = \alpha Df(\hat{x}_0)$$

3. *$f \cdot g$ es derivable en $\hat{x}_0$ y además*

$$D(f \cdot g)(\hat{x}_0) = [f_1(\hat{x}_0) \cdots f_m(\hat{x}_0)]\, Dg(\hat{x}_0) + [g_1(\hat{x}_0) \cdots g_m(\hat{x}_0)]\, Df(\hat{x}_0),$$

en donde $f_1, \ldots, f_m$ y $g_1, \ldots, g_m$ son funciones coordenadas de f y g, respectivamente.

5.2.1. Breve comentario sobre funciones coordenadas

Si el lector revisa con cuidado los conceptos de límite, continuidad, continuidad uniforme y derivabilidad de una función f de $\mathbb{R}^n$ en $\mathbb{R}^m$, todos ellos han sido definidos sin tener que recurrir a algún sistema coordenado, del dominio o del contradominio de la función. Los sistemas coordenados son necesarios e importantes para realizar cálculos específicos y por esta razón hemos probado resultados que expresan estos conceptos en términos de funciones coordenadas.

Así como en el capítulo anterior hicimos énfasis en que la variable $\hat{x} \in \mathbb{R}^n$ de nuestra función f se puede describir en términos de diferentes coordenadas, dependiendo de la base ortonormal de $\mathbb{R}^n$ en la que lo estemos representando, de igual manera las funciones coordenadas asociadas a f también dependen de la base ortonormal de $\mathbb{R}^m$ con la que se esté trabajando. En efecto, si $f_1, \ldots, f_m$ son las funciones coordenadas de f para una cierta base ortonormal $\{\tilde{e}_1, \ldots, \tilde{e}_m\}$ de $\mathbb{R}^m$ (lo que significa que $f_i(\hat{x})$ es la $j-$ésima coordenada del vector $f(\hat{x})$ en esta base), es decir que

$$f(\hat{x}) = f_1(\hat{x})\hat{e}_1 + \cdots + f_m(\hat{x})\hat{e}_m$$

y $\{\tilde{e}'_1, \ldots, \tilde{e}'_m\}$ es otra base ortonormal de $\mathbb{R}^m$, las funciones coordenadas de f en esta nueva base, que denotaremos por $\tilde{f}_1, \ldots, \tilde{f}_m$, se podrán obtener a partir de las funciones coordenadas $f_1, \ldots, f_m$ a través de una matriz ortonormal $\tilde{M} \in M_{m\times m}(\mathbb{R})$, matriz que se obtiene de la misma forma en que obtuvimos la matriz de cambio de coordenadas M de la identidad 4.2 del capítulo 4.

Esto es, si

$$\begin{aligned}\tilde{e}_j &= \left(b_1^{(j)}, \ldots, b_m^{(j)}\right)\\ &= b_1^{(j)}\tilde{e}'_1 + \cdots + b_m^{(j)}\tilde{e}'_m\end{aligned}$$

para cada $j \in \{1, \ldots, m\}$, entonces

$$f(\hat{x}) = f_1(\hat{x})\tilde{e}_1 + \cdots + f_m(\hat{x})\tilde{e}_m$$

$$
\begin{aligned}
&= f_1(\hat{x})\left(b_1^{(1)}\tilde{e}'_1 + \cdots + b_m^{(1)}\tilde{e}'_m\right) + \\
&\vdots \\
&+ f_m(\hat{x})\left(b_1^{(m)}\tilde{e}'_1 + \cdots + b_m^{(m)}\tilde{e}'_m\right) \\
&= \left(b_1^{(1)}, \ldots, b_1^{(m)}\right) \cdot (f_1(\hat{x}), \ldots, f_m(\hat{x}))\tilde{e}'_1 + \\
&\vdots \\
&+ \left(b_m^{(1)}, \ldots, b_m^{(m)}\right) \cdot (f_1(\hat{x}), \ldots, f_m(\hat{x}))\tilde{e}'_m,
\end{aligned}
$$

de donde concluimos que la $j-$ésima coordenada del vector $f(\hat{x})$ en la base $\{\tilde{e}'_1, \ldots, \tilde{e}'_m\}$, que en el párrafo anterior dijimos que denotaríamos por $\tilde{f}_j(\hat{x})$, estará dada por

$$
\tilde{f}_j(\hat{x}) = b_j^{(1)} f_1(\hat{x}) + \cdots + b_j^{(m)} f_m(\hat{x}) \tag{5.3}
$$

para cada $j \in \{1, \ldots, m\}$.

Equivalentemente, pero escrito en forma matricial, se tiene que

$$
\begin{bmatrix} \tilde{f}_1(\hat{x}) & \cdots & \tilde{f}_m(\hat{x}) \end{bmatrix} = \begin{bmatrix} f_1(\hat{x}) & \cdots & f_m(\hat{x}) \end{bmatrix} \begin{bmatrix} b_1^{(1)} & \cdots & b_m^{(1)} \\ \vdots & \ddots & \vdots \\ b_1^{(m)} & \cdots & b_m^{(m)} \end{bmatrix},
$$

en donde la matriz

$$
\tilde{M} = \begin{bmatrix} b_1^{(1)} & \cdots & b_m^{(1)} \\ \vdots & \ddots & \vdots \\ b_1^{(m)} & \cdots & b_m^{(m)} \end{bmatrix}
$$

es una matriz ortonormal.

Lo relevante de la identidad 5.3 es que establece la forma sencilla en que se relacionan las diferentes funciones coordenadas que expresan (en diferentes bases ortonormales) a la misma función. Es decir, si una función f de

$\mathbb{R}^n$ en $\mathbb{R}^m$ tiene funciones coordenadas $f_1, \ldots, f_m$ en una cierta base ortonormal $\{\tilde{e}_1, \ldots, \tilde{e}_m\}$ de $\mathbb{R}^m$, y tomamos otra base ortonormal $\{\tilde{e}'_1, \ldots, \tilde{e}'_m\}$ de este mismo conjunto, sus funciones coordenadas $\tilde{f}_1, \ldots, \tilde{f}_m$ en esta nueva base se obtienen como una combinación lineal de las primeras. Esta relación resulta de particular importancia para todos los conceptos que mencionamos al inicio de esta subsección, pues confirma el hecho de que todos ellos son independientes del sistema coordenado de $\mathbb{R}^m$ que se esté usando para representar a f.

Para el caso del concepto de derivada, la identidad 5.3 resulta particularmente importante, pues si tenemos una expresión para la derivada $Df(\hat{x}_0)$ en términos de unas funciones coordenadas $f_1, \ldots, f_m$, de esta identidad podemos deducir cuál será una expresión de esta derivada, pero ahora en términos de las funciones coordenadas $\tilde{f}_1, \ldots, \tilde{f}_m$. En efecto, si escribimos a $Df(\hat{x}_0)$ como $D\left(\tilde{f}_1, \ldots, \tilde{f}_m\right)(\hat{x}_0)$ para enfatizar que estamos escribiendo a f en términos de las funciones coordenadas $\tilde{f}_1, \ldots, \tilde{f}_m$, de acuerdo con la identidad 5.2 se debe tener que

$$\begin{aligned} Df(\hat{x}_0) &= D\left(\tilde{f}_1, \ldots, \tilde{f}_m\right)(\hat{x}_0) \\ &= \begin{bmatrix} \frac{\partial \tilde{f}_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial \tilde{f}_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{f}_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial \tilde{f}_m}{\partial x_n}(\hat{x}_0) \end{bmatrix}, \end{aligned}$$

de tal forma que, de la identidad 5.3 y los incisos 1. y 2. de la proposición 4.6 del capítulo 4, concluimos que, para cada $i \in \{1, \ldots, n\}$ y cada $j \in \{1, \ldots, m\}$, se tiene que

$$\begin{aligned} \frac{\partial \tilde{f}_j}{\partial x_i}(\hat{x}_0) &= b_j^{(1)} \frac{f_1}{\partial x_i}(\hat{x}_0) + \cdots + b_j^{(m)} \frac{f_m}{\partial x_i}(\hat{x}_0) \\ &= \left(b_j^{(1)}, \ldots, b_j^{(m)}\right) \cdot \left(\frac{f_1}{\partial x_i}(\hat{x}_0), \ldots, \frac{f_m}{\partial x_i}(\hat{x}_0)\right), \end{aligned}$$

de donde obtenemos que

$$\begin{aligned} D\left(\tilde{f}_1,\ldots,\tilde{f}_m\right)(\hat{x}_0) &= \begin{bmatrix} \frac{\partial \tilde{f}_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial \tilde{f}_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{f}_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial \tilde{f}_m}{\partial x_n}(\hat{x}_0) \end{bmatrix} \\ &= \begin{bmatrix} b_1^{(1)} & \cdots & b_1^{(m)} \\ \vdots & \ddots & \vdots \\ b_m^{(1)} & \cdots & b_m^{(m)} \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\hat{x}_0) \end{bmatrix} \\ &= \tilde{M}^t D\left(f_1,\ldots,f_m\right)(\hat{x}_0). \end{aligned}$$

5.3. La regla de la cadena

Para terminar de revisar las propiedades más relevantes de la derivada de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$, formularemos un resultado muy importante en el cual se establecen las condiciones para asegurar la derivabilidad de la composición de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$ con una de $\mathbb{R}^m$ en $\mathbb{R}^k$, y la fórmula que nos permite calcular la derivada de esta composición. Esta será la versión más general de *la regla de la cadena* que daremos en este texto, y para probarla podríamos hacer uso de la proposición 5.1 de la siguiente manera: si $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ y $g : V \subset \mathbb{R}^m \to \mathbb{R}^k$, con $g_1,\ldots,g_m$ funciones coordenadas de g, son tales que $f(U) \subset V$, se tiene que la composición $g \circ f$ está bien definida y que las funciones $g_i \circ f$, para $i \in \{1,\ldots,m\}$, son funciones coordenadas de $g \circ f$.

De esta forma, y justo por la proposición 5.1, para probar la derivabilidad de $g \circ f$ bastaría con probar la derivabilidad de cada $g_i \circ f$. Sin embargo, la prueba de la regla de la cadena para la composición $g \circ f$, suponiendo que g es una función de $\mathbb{R}^m$ en $\mathbb{R}$, es casi igual a la prueba que se puede hacer suponiendo que g es una función de $\mathbb{R}^m$ en $\mathbb{R}^k$, en virtud de lo cual optaremos por esta última.

Proposición 5.6 (Regla de la cadena) *Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $g : V \subset \mathbb{R}^m \to \mathbb{R}^k$ y $\hat{x}_0 \in U$ tales que $f(U) \subset V$. Si f es derivable en $\hat{x}_0$ y g*

es derivable en $\hat{y}_0 = f(\hat{x}_0)$*, entonces* $g \circ f$ *es derivable en* $\hat{x}_0$ *y además se tiene que*

$$D(g \circ f)(\hat{x}_0) = Dg(f(\hat{x}_0)) \circ Df(\hat{x}_0).$$

O equivalentemente, si pensamos a las derivadas $D(g \circ f)(\hat{x}_0)$*,* $Dg(f(\hat{x}_0))$ *y* $Df(\hat{x}_0)$ *como matrices, escribimos que*

$$D(g \circ f)(\hat{x}_0) = Dg(f(\hat{x}_0))Df(\hat{x}_0).$$

Demostración. La idea detrás de la prueba de este resultado es completamente análoga a la seguida en la prueba de la proposicón 4.16 del capítulo 4, en virtud de lo cual haremos con menos detalles los pasos que seguiremos en este caso.

Como hicimos en esa proposición, será necesario introducir una función auxiliar

$$\varphi : U \subset \mathbb{R}^n \to \mathbb{R}^k$$

definida de la siguiente forma:

$$\varphi(\hat{x}) = \begin{cases} \frac{g(f(\hat{x})) - g(f(\hat{x}_0)) - Dg(f(\hat{x}_0))(f(\hat{x}) - f(\hat{x}_0))}{\|f(\hat{x}) - f(\hat{x}_0)\|} & \text{si } f(\hat{x}) - f(\hat{x}_0) \neq \hat{0} \\ \hat{0} & \text{si } f(\hat{x}) - f(\hat{x}_0) = \hat{0} \end{cases}$$

Como el lector podrá verificar muy fácilmente, se tiene que

$$\begin{aligned} &\frac{g(f(\hat{x})) - g(f(\hat{x}_0)) - Dg(f(\hat{x}_0))\left(Df(\hat{x}_0)\left(\hat{x} - \hat{x}_0\right)\right)}{\|\hat{x} - \hat{x}_0\|} \\ &= \varphi(\hat{x})\frac{\|f(\hat{x}) - f(\hat{x}_0)\|}{\|\hat{x} - \hat{x}_0\|} + \frac{Dg(f(\hat{x}_0))\left(f(\hat{x}) - f(\hat{x}_0)\right)}{\|\hat{x} - \hat{x}_0\|} \\ &\quad - \frac{Dg(f(\hat{x}_0))\left(Df(\hat{x}_0)\left(\hat{x} - \hat{x}_0\right)\right)}{\|\hat{x} - \hat{x}_0\|} \end{aligned}$$

$$= \varphi(\hat{x})\frac{\|f(\hat{x}) - f(\hat{x}_0)\|}{\|\hat{x} - \hat{x}_0\|} + Dg(f(\hat{x}_0))\left(\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)\left(\hat{x} - \hat{x}_0\right)}{\|\hat{x} - \hat{x}_0\|}\right)$$

para toda $\hat{x} \in U$, $\hat{x} \neq \hat{x}_0$.

Ahora, del hecho de que f es derivable en $\hat{x}_0$, y por los problemas 5 y 4 de este capítulo, obtenemos las dos siguientes conclusiones: una, que la función φ es continua en $\hat{x}_0$, y dos, que la expresión

$$\frac{\|f(\hat{x}) - f(\hat{x}_0)\|}{\|\hat{x} - \hat{x}_0\|}$$

está acotada en una vecindad (agujerada) de $\hat{x}_0$, de tal forma que

$$\lim_{\hat{x}\to\hat{x}_0} \varphi(\hat{x}) \frac{\|f(\hat{x}) - f(\hat{x}_0)\|}{\|\hat{x} - \hat{x}_0\|} = \hat{0}.$$

Por otra parte, dado que f es derivable en $\hat{x}_0$, sabemos que

$$\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)\,(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} = \hat{0}$$

y como toda función lineal es continua, tenemos que

$$\begin{aligned}
&\lim_{\hat{x}\to\hat{x}_0} Dg(f(\hat{x}_0)) \left(\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)\,(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \right) \\
&= Dg(f(\hat{x}_0)) \left(\lim_{\hat{x}\to\hat{x}_0} \frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)\,(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \right) \\
&= Dg(f(\hat{x}_0)) \left(\hat{0}\right) \\
&= \hat{0}.
\end{aligned}$$

De esta forma,

$$\begin{aligned}
&\lim_{\hat{x}\to\hat{x}_0} \frac{g(f(\hat{x})) - g(f(\hat{x}_0)) - Dg(f(\hat{x}_0))\,(Df(\hat{x}_0)\,(\hat{x} - \hat{x}_0))}{\|\hat{x} - \hat{x}_0\|} \\
&= \lim_{\hat{x}\to\hat{x}_0} \left[\varphi(\hat{x}) \frac{\|f(\hat{x}) - f(\hat{x}_0)\|}{\|\hat{x} - \hat{x}_0\|} \right. \\
&\qquad \left. + Dg(f(\hat{x}_0)) \left(\frac{f(\hat{x}) - f(\hat{x}_0) - Df(\hat{x}_0)\,(\hat{x} - \hat{x}_0)}{\|\hat{x} - \hat{x}_0\|} \right) \right] \\
&= \hat{0},
\end{aligned}$$

lo que prueba que $g \circ f$ es derivable en $\hat{x}_0$ y que su derivada es $Dg(f(\hat{x}_0)) \circ Df(\hat{x}_0)$. ■

5.3.1. Cambio de coordenadas y regla de la cadena

Desde el capítulo 1, a lo largo de todo este texto se ha venido insistiendo en ver a $\mathbb{R}^n$ como el "representante por excelencia" (o "prototipo") de los espacios vectoriales de dimensión n sobre los números reales. Por esta misma razón, se ha enfatizado la posibilidad de que los elementos de $\mathbb{R}^n$ se pueden describir por medio de diferentes sistemas coordenados (lo que no deja de ser un poco extraño, pues los elementos de $\mathbb{R}^n$ están definidos en términos de n-adas). Más aún, con base en la representación geométrica de $\mathbb{R}^2$ y $\mathbb{R}^3$ se han incluido otros sistemas coordenados no cartesianos, como lo son los sistemas coordenados cilíndrico y esférico (en $\mathbb{R}^3$), y polar (en $\mathbb{R}^2$).

Con base en lo anterior, se ha puesto particular interés en mostrar que el concepto de derivada es independiente de estos sistemas coordenados, pero al mismo tiempo se ha visto cómo se expresa en éstos, así como la manera de "pasar" de una expresión a otra cuando "pasamos" de un sistema coordenado a otro (ejemplo de esto son las identidades 4.18, 4.19, 4.33 y 4.34). Lo que ahora deseamos mostrar es que estas mismas "identidades" se pueden obtener usando la regla de la cadena que acabamos de probar. Por ejemplo, supongamos que $\{\hat{e}_1, \ldots, \hat{e}_n\}$ y $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$ son dos bases ortonormales de $\mathbb{R}^n$, y que $x_1, \ldots, x_n$ y $x'_1, \ldots, x'_n$ denotan las coordenadas de un mismo punto $\hat{x} \in \mathbb{R}^n$ en cada una de estas bases, respectivamente. Si

$$\hat{e}'_i = a_1^{(i)}\hat{e}_1 + \cdots + a_n^{(i)}\hat{e}_n$$

para cada $i \in \{1, \ldots, n\}$, en el capítulo 4 dedujimos que si $(x'_1, \ldots, x'_n)$ son las coordenadas de un punto $\hat{x} \in \mathbb{R}^n$ en la base $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$, entonces las coordenadas $(x_1, \ldots, x_n)$ del mismo punto en la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$ están dadas por

$$\left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right] = \left[\begin{array}{ccc} x'_1 & \cdots & x'_n \end{array}\right] \left[\begin{array}{ccc} a_1^{(1)} & \cdots & a_n^{(1)} \\ \vdots & \ddots & \vdots \\ a_1^{(n)} & \cdots & a_n^{(n)} \end{array}\right].$$

De esta forma, si definimos $g : \mathbb{R}^n \to \mathbb{R}^n$ como

$$\begin{aligned} g(\hat{x}) &= g(x'_1, \ldots, x'_n) \\ &= \left(x'_1 a_1^{(1)} + \cdots + x'_n a_1^{(n)}, \ldots, x'_1 a_n^{(1)} + \cdots + x'_n a_n^{(n)}\right), \end{aligned}$$

se tendrá que g no es más que la función identidad, sólo que estamos expresando a los elementos de su dominio en la base $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$, mientras que a los de su contradominio los estamos expresando en la base $\{\hat{e}_1, \ldots, \hat{e}_n\}$. Este tipo de funciones son conocidas como *funciones de cambio de coordenadas*.

Dado que g es la función identidad, sin duda g es derivable en todos los puntos de $\mathbb{R}^n$. Por otra parte, de acuerdo con la identidad 5.1, tenemos que la matriz jacobiana de g (es decir $Dg(\hat{x})$), expresada con respecto a las bases $\{\hat{e}_1, \ldots, \hat{e}_n\}$ y $\{\hat{e}'_1, \ldots, \hat{e}'_n\}$, está dada por

$$Dg(\hat{x}) = \left[\begin{array}{ccc} a_1^{(1)} & \cdots & a_1^{(n)} \\ \vdots & \ddots & \vdots \\ a_n^{(1)} & \cdots & a_n^{(n)} \end{array}\right]$$

para toda $\hat{x} \in \mathbb{R}^n$, que coincide con la matriz transpuesta de la matriz de cambio de coordenadas que obtuvimos anteriormente.

Ahora, si $f : U \subset \mathbb{R}^n \to \mathbb{R}$ es una función derivable que está expresada en términos de las coordenadas $x_1, \ldots, x_n$, dado que g es la función identidad, que la composición $f \circ g$ está bien definida y que en sentido estricto $(f \circ g)(\hat{x}) = f(\hat{x})$ para toda $\hat{x} \in U$ (identidad que no es equivalente a escribir que $f(g(x'_1, \ldots, x'_n)) = f(x'_1, \ldots, x'_n)$, puesto que f no depende de las coordenadas $x'_1, \ldots, x'_n$), podemos decir que $f \circ g$ es la misma función f, pero expresada en términos de las coordenadas $x'_1, \ldots, x'_n$. Este hecho es lo que justifica que nos podamos tomar la libertad de escribir que

$$\frac{\partial\,(f \circ g)}{\partial x'_i}(\hat{x}) = \frac{\partial f}{\partial x'_i}(\hat{x})$$

para cada $i \in \{1, \ldots, n\}$ y cada $\hat{x} \in U$, aún cuando f no dependa directamente de las coordenadas $x'_1, \ldots, x'_n$.

Por otra parte, de acuerdo con la identidad (matricial) de la regla de la cadena (proposición 5.6), se tiene que

$$\begin{aligned} D(f \circ g)(\hat{x}) &= Df(g(\hat{x}))Dg(\hat{x}) \\ &= Df(\hat{x})Dg(\hat{x}). \end{aligned} \tag{5.4}$$

Es decir, que

$$\begin{aligned} &\left[\frac{\partial f}{\partial x'_i}(\hat{x}) \cdots \frac{\partial f}{\partial x'_n}(\hat{x}) \right] \\ &\qquad = \left[\begin{array}{ccc} \frac{\partial (f \circ g)}{\partial x'_1}(\hat{x}) & \cdots & \frac{\partial (f \circ g)}{\partial x'_n}(\hat{x}) \end{array} \right] \\ &\qquad = \left[\begin{array}{ccc} \frac{\partial f}{\partial x_i}(\hat{x}) & \cdots & \frac{\partial f}{\partial x_n}(\hat{x}) \end{array} \right] \left[\begin{array}{ccc} a_1^{(1)} & \cdots & a_1^{(n)} \\ \vdots & \ddots & \vdots \\ a_n^{(1)} & \cdots & a_n^{(n)} \end{array} \right] \end{aligned} \tag{5.5}$$

o equivalentemente, que para cada $i \in \{1, \ldots, n\}$ se cumple que

$$\begin{aligned} \frac{\partial f}{\partial x'_i}(\hat{x}) &= a_1^{(i)} \frac{\partial f}{\partial x_1}(\hat{x}) + \cdots + a_n^{(i)} \frac{\partial f}{\partial x_n}(\hat{x}) \\ &= \sum_{j=1}^{n} a_j^{(i)} \frac{\partial f}{\partial x_j}(\hat{x}). \end{aligned} \tag{5.6}$$

Como seguramente el lector reconocerá, la identidad 5.5 es la misma que obtuvimos en el capítulo 4 (identidad 4.18).

Observación 5.2 *Es relevante hacer notar que las identidades 5.4, 5.5 y 5.6 conllevan un cierto abuso de notación, el cual puede ser "inofensivo" sólo si dichas identidades se escriben en términos de $\hat{x}$ y no de sus correspondientes coordenadas $x'_1, \ldots, x'_n$. Si deseamos poner a la identidad 5.6 en términos de estas coordenadas, lo correcto es escribir que*

$$\begin{aligned}
&\frac{\partial\,(f\circ g)}{\partial x_i'}(x_1',\ldots,x_n')\\
&= a_1^{(i)}\frac{\partial f}{\partial x_1}(g\left(x_1',\ldots,x_n'\right))+\cdots+a_n^{(i)}\frac{\partial f}{\partial x_n}(g\left(x_1',\ldots,x_n'\right))\\
&= a_1^{(i)}\left(\frac{\partial f}{\partial x_1}\circ g\right)(x_1',\ldots,x_n')+\cdots+a_n^{(i)}\left(\frac{\partial f}{\partial x_n}\circ g\right)(x_1',\ldots,x_n')\,.
\end{aligned}$$

Es muy importante tener presente lo anterior, sobre todo si se desea calcular derivadas parciales de orden superior de la función $f\circ g$.

Para concluir esta subsección, lo que ahora queremos mostrar es que se puede proceder de forma análoga al caso anterior y obtener identidades equivalentes, aún cuando los sistemas involucrados no sean cartesianos.

Por ejemplo, consideremos la función $g:\mathbb{R}^2\to\mathbb{R}^2$ dada por

$$g(\rho,\theta)=(\rho\cos(\theta),\rho\,\mathrm{sen}(\theta))\,,$$

la que sin duda se puede considerar como la *función de cambio de coordenadas polares a coordenadas cartesianas* para puntos en $\mathbb{R}^2$, y supongamos que $f:\mathbb{R}^2\to\mathbb{R}$ es una función derivable que está expresada en términos de las coordenadas cartesianas x y y.

A diferencia del caso cartesiano, esta función de cambio de coordenadas es tal que su derivada no es constante, además de que la matriz que la representa (tomando la base canónica de $\mathbb{R}^2$, tanto en el dominio como en el contradominio) no es una matriz ortonormal para toda pareja (ρ,θ). Sin embargo, la aplicación de la regla de la cadena nos permite deducir algunas identidades que resultarán muy útiles.

En efecto, por la regla de la cadena sabemos que

$$D(f\circ g)(\hat{x})=Df(g(\hat{x}))Dg(\hat{x})$$

o equivalentemente, que

$$\nabla(f\circ g)(\hat{x})=\nabla f(g(\hat{x}))Dg(\hat{x}).$$

Si ahora, como dijimos anteriormente, escribimos que

$$\begin{aligned} Dg(\hat{x}) &= \begin{bmatrix} \frac{\partial}{\partial \rho}\left(\rho\cos(\theta)\right) & \frac{\partial}{\partial \theta}\left(\rho\cos(\theta)\right) \\ \frac{\partial}{\partial \rho}\left(\rho\,\mathrm{sen}(\theta)\right) & \frac{\partial}{\partial \theta}\left(\rho\,\mathrm{sen}(\theta)\right) \end{bmatrix} \\ &= \begin{bmatrix} \cos(\theta) & -\rho\,\mathrm{sen}(\theta) \\ \mathrm{sen}(\theta) & \rho\cos(\theta) \end{bmatrix} \end{aligned}$$

y

$$\begin{aligned} \nabla(f\circ g)(\hat{x}) &= \begin{bmatrix} \frac{\partial(f\circ g)}{\partial \rho}(\hat{x}) & \frac{\partial(f\circ g)}{\partial \theta}(\hat{x}) \end{bmatrix} \\ \nabla f(g(\hat{x})) &= \begin{bmatrix} \frac{\partial f}{\partial x}(g(\hat{x})) & \frac{\partial f}{\partial y}(g(\hat{x})) \end{bmatrix}, \end{aligned}$$

concluimos que

$$\begin{aligned} \begin{bmatrix} \frac{\partial(f\circ g)}{\partial \rho}(\hat{x}) & \frac{\partial(f\circ g)}{\partial \theta}(\hat{x}) \end{bmatrix} &= \nabla(f\circ g)(\hat{x}) \\ &= \nabla f(g(\hat{x}))Dg(\hat{x}) \\ &= \begin{bmatrix} \frac{\partial f}{\partial x}\left(g(\hat{x})\right) & \frac{\partial f}{\partial y}\left(g(\hat{x})\right) \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\rho\,\mathrm{sen}(\theta) \\ \mathrm{sen}(\theta) & \rho\cos(\theta) \end{bmatrix}. \end{aligned}$$

De esta última identidad, si definimos $\hat{e}_\rho(\hat{x}) = \cos(\theta)\hat{e}_1 + \mathrm{sen}(\theta)\hat{e}_2$ y $\hat{e}_\theta(\hat{x}) = -\mathrm{sen}(\theta)\hat{e}_1 + \cos(\theta)\hat{e}_2$, obtenemos que

$$\begin{aligned} \frac{\partial(f\circ g)}{\partial \rho}(\hat{x}) &= \cos(\theta)\frac{\partial f}{\partial x}\left(g(\hat{x})\right) + \mathrm{sen}(\theta)\frac{\partial f}{\partial y}\left(g(\hat{x})\right) \\ &= \nabla f(g(\hat{x}))\cdot \hat{e}_\rho(\hat{x}) \end{aligned} \tag{5.7}$$

y

$$\begin{aligned} \frac{\partial(f\circ g)}{\partial \theta}(\hat{x}) &= -\rho\,\mathrm{sen}(\theta)\frac{\partial f}{\partial x}\left(g(\hat{x})\right) + \rho\cos(\theta)\frac{\partial f}{\partial y}\left(g(\hat{x})\right) \\ &= \rho\nabla f(g(\hat{x}))\cdot \hat{e}_\theta(\hat{x}), \end{aligned} \tag{5.8}$$

o equivalentemente, si $\rho \neq 0$, que

$$\nabla f(g(\hat{x}))\cdot \hat{e}_\rho(\hat{x}) = \frac{\partial(f\circ g)}{\partial \rho}(\hat{x})$$

y

$$\nabla f(g(\hat{x})) \cdot \hat{e}_\theta(\hat{x}) = \frac{1}{\rho}\frac{\partial(f \circ g)}{\partial \theta}(\hat{x}).$$

Dado que los vectores $\hat{e}_\rho(\hat{x})$ y $\hat{e}_\theta(\hat{x})$ forman una base ortonormal de $\mathbb{R}^2$ (la cual depende del punto $\hat{x}$), las últimas dos identidades establecen que las coordenadas de $\nabla f(g(\hat{x}))$ en la base $\{\hat{e}_\rho(\hat{x}), \hat{e}_\theta(\hat{x})\}$ están dadas por $\frac{\partial(f\circ g)}{\partial \rho}(\hat{x})$ y $\frac{1}{\rho}\frac{\partial(f\circ g)}{\partial \theta}(\hat{x})$, es decir, que

$$\nabla f(g(\hat{x})) = \frac{\partial(f \circ g)}{\partial \rho}(\hat{x})\hat{e}_\rho(\hat{x}) + \frac{1}{\rho}\frac{\partial(f \circ g)}{\partial \theta}(\hat{x})\hat{e}_\theta(\hat{x}).$$

Si no usamos algún tipo de coordenadas para representar al punto $\hat{x}$, entonces podemos asumir que $g(\hat{x}) = \hat{x}$ y así tomarnos la libertad de escribir que $f \circ g = f$. De esta identidad, tendremos que

$$\frac{\partial(f \circ g)}{\partial \rho}(\hat{x}) = \frac{\partial f}{\partial \rho}(\hat{x}) \qquad \text{y} \qquad \frac{\partial(f \circ g)}{\partial \theta}(\hat{x}) = \frac{\partial f}{\partial \theta}(\hat{x})$$

y concluimos que

$$\nabla f\,(\hat{x}) = \frac{\partial f}{\partial \rho}(\hat{x})\hat{e}_\rho(\hat{x}) + \frac{1}{\rho}\frac{\partial f}{\partial \theta}(\hat{x})\hat{e}_\theta(\hat{x}).$$

Como el lector podrá verficar fácilmente, esta últma identidad coincide con la identidad de matrices 4.34 del capítulo 4.

Aun corriendo el riesgo de parecer repetitivos, es importante insistir en que las identidades

$$\frac{\partial f}{\partial \rho}(\hat{x}) = \cos(\theta)\frac{\partial f}{\partial x}(\hat{x}) + \operatorname{sen}(\theta)\frac{\partial f}{\partial y}(\hat{x})$$

y

$$\frac{\partial f}{\partial \theta}(\hat{x}) = -\rho\operatorname{sen}(\theta)\frac{\partial f}{\partial x}(\hat{x}) + \rho\cos(\theta)\frac{\partial f}{\partial y}(\hat{x})$$

(que se obtienen a partir de las identidades 5.7 y 5.8 y de escribir que $f \circ g = f$) contienen un cierto abuso de notación. Si se quieren escribir en términos de las coordenadas polares (ρ, θ) del punto $\hat{x}$, lo correcto será escribir que

$$\frac{\partial\,(f\circ g)}{\partial\rho}(\rho,\theta)=\cos(\theta)\frac{\partial f}{\partial x}(g(\rho,\theta))+\operatorname{sen}(\theta)\frac{\partial f}{\partial y}(g(\rho,\theta))$$
$$=\cos(\theta)\left(\frac{\partial f}{\partial x}\circ g\right)(\rho,\theta)+\operatorname{sen}(\theta)\left(\frac{\partial f}{\partial y}\circ g\right)(\rho,\theta)$$

y

$$\frac{\partial\,(f\circ g)}{\partial\theta}(\rho,\theta)=-\rho\operatorname{sen}(\theta)\frac{\partial f}{\partial x}(g(\rho,\theta))+\rho\cos(\theta)\frac{\partial f}{\partial y}(g(\rho,\theta))$$
$$=-\rho\operatorname{sen}(\theta)\left(\frac{\partial f}{\partial x}\circ g\right)(\rho,\theta)+\rho\cos(\theta)\left(\frac{\partial f}{\partial y}\circ g\right)(\rho,\theta).$$

Como ya se mencionó antes, estas identidades son las que hay que considerar si se desea calcular las derivadas parciales de orden superior (con respecto a las variables ρ o θ) de la función $f\circ g$.

5.4. El teorema de la función implícita

Una vez que hemos desarrollado las herramientas básicas relacionadas con el concepto de derivada para funciones de $\mathbb{R}^n$ en $\mathbb{R}^m$, estamos en condiciones de abordar uno de los teoremas más importantes del cálculo diferencial de varias variables: el Teorema de la Función Implícita.

5.4.1. El caso lineal

Con el fin de motivar y deducir el contenido de este teorema, empezaremos por analizar algunas propiedades y características de las soluciones de los sistemas de ecuaciones lineales en $\mathbb{R}^n$ que tienen la particularidad de tener menos ecuaciones que incógnitas, es decir, sistemas de la forma

$$\begin{aligned}
a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n+b_1&=0\\
a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n+b_2&=0\\
&\vdots\\
a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n+b_m&=0,
\end{aligned}\tag{5.9}$$

en donde $m<n$.

Sin duda el caso más sencillo de este tipo de sistemas de ecuaciones lo tenemos en $\mathbb{R}^2$, y se trata de un sistema que consta de una sola ecuación de la forma

$$ax + by + c = 0$$

Como seguramente recordará el lector, si $a^2+b^2 > 0$ la ecuación anterior representa una recta en el plano, y si de obtener sus soluciones se trata, es fácil probar que si $a \neq 0$, entonces éstas son de la forma

$$\left(\frac{-by - c}{a}, y\right)$$

para cualquier $y \in \mathbb{R}$, y si $b \neq 0$, entonces todas las soluciones las podemos escribir como las parejas

$$\left(x, \frac{-ax - c}{b}\right).$$

Es decir, si el coeficiente de la incógnita x es diferente de cero ($a \neq 0$), esta incógnita se puede poner en función de la otra incógnita y ($x = h(y) = (-by - c)/a$) y todas las soluciones de la ecuación son de la forma $(h(y), y)$, con $y \in \mathbb{R}$.

Sucede lo análogo si el coeficiente de la incógnita y es diferente de cero ($b \neq 0$); en este caso y se puede poner en función de x ($y = h(x) = (-ax - c)/b$) y todas las soluciones de la ecuación son de la forma $(x, h(x))$, con $x \in \mathbb{R}$.

Para sistemas de ecuaciones en $\mathbb{R}^3$, la situación empieza a ponerse un poco más interesante. El primer caso que se puede tener es el de una sóla ecuación de la forma

$$ax + by + cz + d = 0 \tag{5.10}$$

(la cual geométricamente representa a un plano si $a^2 + b^2 + c^2 > 0$), y la obtención de sus soluciones es muy similar al caso anterior. En efecto, nótese que si el coeficiente de la incógnita x es diferente de cero ($a \neq 0$), entonces podemos poner a x en función de las incógnitas y y z como

$$\begin{aligned} x &= h(y, z) \\ &= \frac{-(by + cz + d)}{a} \end{aligned}$$

y todas las ternas de la forma

$$(h(y,z),y,z)=\left(\frac{-(by+cz+d)}{a},y,z\right),$$

con $(y,z)\in\mathbb{R}^2$, son las soluciones de la ecuación 5.10. Como el lector puede constatar fácilmente, se tiene una situación análoga si $b\neq 0$ o $c\neq 0$.

El caso que se empieza a poner aún más interesante es cuando tenemos un sistema de dos ecuaciones de la forma

$$\begin{aligned} a_{11}x+a_{12}y+a_{13}z+b_1&=0\\ a_{21}x+a_{22}y+a_{23}z+b_2&=0. \end{aligned} \tag{5.11}$$

Lo primero que haremos, será recordar que en un sistema de este tipo se pueden presentar las siguientes situaciones:

1. que una de las ecuaciones sea "múltiplo" de la otra, es decir, que una de ellas se puede obtener de la otra multiplicando por un escalar, lo que significa que en realidad nuestro sistema sólo consta de una ecuación, caso que ya analizamos

2. que los planos que representan cada una de las ecuaciones, sean dos planos distintos pero paralelos, es decir, que los vectores normales a cada uno de ellos ((a_{11},a_{12},a_{13}) y (a_{21},a_{22},a_{23})), sean paralelos; en este caso no existen soluciones y no hay nada más que se pueda hacer

3. la tercera y última posibilidad es justo cuando los vectores (a_{11},a_{12},a_{13}) y (a_{21},a_{22},a_{23}) no son paralelos, y por lo tanto el conjunto de soluciones del sistema está formado por todos los puntos de la recta en la que se intersectan ambos planos.

Dado que los dos primeros casos ya están resueltos, analizaremos el tercero. Del hecho de que los vectores (a_{11},a_{12},a_{13}) y (a_{21},a_{22},a_{23}) no sean paralelos podemos concluir entonces que el producto cruz de éstos no es el vector $\hat{0}$, es decir que

$$\begin{aligned}(a_{11}, a_{12}, a_{13}) \times (a_{21}, a_{22}, a_{23}) & \\ = (a_{12}a_{23} - a_{13}a_{22}, a_{13}a_{21} - a_{11}a_{23}, a_{11}a_{22} - a_{12}a_{21}) & \quad (5.12)\\ \neq \hat{0} &\end{aligned}$$

y por lo tanto tenemos tres posibilidades: $a_{12}a_{23} - a_{13}a_{22} \neq 0$, $a_{13}a_{21} - a_{11}a_{23} \neq 0$ o $a_{11}a_{22} - a_{12}a_{21} \neq 0$, posibilidades que a continuación nos disponemos a analizar.

Lo primero que habría que destacar es que cada una de las coordenadas del vector dado por 5.12, resulta ser el determinante de la matriz formada por los coeficientes de algunas de las incógnitas de nuestras ecuaciones. En efecto, nótese que $a_{12}a_{23} - a_{13}a_{22}$ es el determinante de la matriz que se obtiene al considerar sólo los coeficientes (de ambas ecuaciones del sistema dado por 5.11) de las incógnitas y y z; $a_{13}a_{21} - a_{11}a_{23}$ es el determinante de la matriz que se obtiene al considerar sólo los coeficientes de las incógnitas x y z, y $a_{11}a_{22} - a_{12}a_{21}$ es el determinante de la matriz que se obtiene al considerar solo los coeficientes de las incógnitas x y y.

Con base en la observación anterior, si por ejemplo se tiene que $a_{12}a_{23} - a_{13}a_{22} \neq 0$, todo parece indicar que lo más adecuado sería reescribir el sistema de ecuaciones 5.11 en la forma

$$\begin{aligned} a_{12}y + a_{13}z &= -a_{11}x - b_1 \\ a_{22}y + a_{23}z &= -a_{21}x - b_2, \end{aligned}$$

o mejor aún, en la forma matricial

$$\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} -(a_{11}x + b_1) \\ -(a_{21}x + b_2) \end{bmatrix}.$$

De manera que, si escribimos

$$M = \begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix},$$

como $\det(M) = a_{12}a_{23} - a_{13}a_{22} \neq 0$, sabemos que M es invertible y por lo tanto tendremos que las incógnitas y y z se pueden poner en función de la incógnita x, como

$$\begin{bmatrix} y \\ z \end{bmatrix} = M^{-1} \begin{bmatrix} -(a_{11}x + b_1) \\ -(a_{21}x + b_2) \end{bmatrix} = \begin{bmatrix} h_1(x) \\ h_2(x) \end{bmatrix}$$

y que las soluciones del sistema 5.11 estarían dadas por las ternas

$$(x, h_1(x), h_2(x)),$$

con $x \in \mathbb{R}$.

Tomando en consideración el análisis anterior, seguramente el lector estará de acuerdo en que, si ahora lo que se tiene es que $a_{13}a_{21} - a_{11}a_{23} \neq 0$, entonces las incógnitas x y z se podrán poner en función de la incógnita y ($x = h_1(y)$ y $z = h_2(y)$) y que las soluciones del sistema estarán dadas por las ternas $(h_1(y), y, h_2(y))$ con $y \in \mathbb{R}$. Y si lo que sucede es que $a_{11}a_{22} - a_{12}a_{21} \neq 0$, entonces las incógnitas x y y se podrán poner en función de la incógnita z ($x = h_1(z)$ y $y = h_2(z)$) y que las soluciones del sistema estarán dadas por las ternas $(h_1(z), h_2(z), z)$, con $z \in \mathbb{R}$.

Sin duda el caso anterior es muy ilustrativo y nos da la pauta para resolver el problema general sobre el cálculo de las soluciones del sistema de ecuaciones dado por 5.9. De esta manera, si por ejemplo se tiene que la matriz de $m \times m$ formada por los coeficientes de las incógnitas $x_1, \ldots, x_m$, dada por

$$M = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix}$$

tiene determinante distinto de cero (y por lo tanto tiene inversa), entonces el sistema 5.9 escrito en términos de la matriz M toma la forma

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} -\left(a_{1(m+1)}x_{m+1} + \cdots + a_{1n}x_n + b_1\right) \\ \vdots \\ -\left(a_{m(m+1)}x_{m+1} + \cdots + a_{mn}x_n + b_m\right) \end{bmatrix}$$

y por lo tanto las incógnitas $x_1, \ldots, x_m$ se podrán poner en función de las incógnitas restantes $x_{m+1}, \ldots, x_n$ como

$$\begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = M^{-1} \begin{bmatrix} -\left(a_{1(m+1)}x_{m+1} + \cdots + a_{1n}x_n + b_1\right) \\ \vdots \\ -\left(a_{m(m+1)}x_{m+1} + \cdots + a_{mn}x_n + b_m\right) \end{bmatrix}.$$

De esta forma, si definimos

$$H = (h_1, \ldots, h_m) : \mathbb{R}^{n-m} \to \mathbb{R}^m$$

como

$$H(x_{m+1}, \ldots, x_n) = M^{-1} \begin{bmatrix} -\left(a_{1(m+1)}x_{m+1} + \cdots + a_{1n}x_n + b_1\right) \\ \vdots \\ -\left(a_{m(m+1)}x_{m+1} + \cdots + a_{mn}x_n + b_m\right) \end{bmatrix},$$

entonces encontramos que existen $h_1, \ldots, h_m : \mathbb{R}^{n-m} \to \mathbb{R}$ tales que

$$x_j = h_j(x_{m+1}, \ldots, x_n)$$

para $j \in \{1, \ldots, m\}$, y que las soluciones del sistema de ecuaciones 5.9 están dadas por las $n-$adas

$$(h_1(x_{m+1}, \ldots, x_n), \ldots, h_m(x_{m+1}, \ldots, x_n), x_{m+1}, \ldots, x_n) \in \mathbb{R}^n$$

con $(x_{m+1}, \ldots, x_n) \in \mathbb{R}^{n-m}$.

En el caso general se tendrá que, si los índices $1 \leq i_1 < i_2 < \cdots < i_m \leq n$, $1 \leq j_{m+1} < j_{m+2} < \cdots < j_n \leq n$, con $\{i_1, \ldots, i_m\} \cap \{j_{m+1}, \ldots, j_n\} = \varnothing$, son tales que la matriz

$$\begin{bmatrix} a_{1i_1} & \cdots & a_{1i_m} \\ \vdots & \ddots & \vdots \\ a_{mi_1} & \cdots & a_{mi_m} \end{bmatrix}$$

es invertible, entonces las incógnitas $x_{i_1}, \ldots, x_{i_m}$ se pueden poner en función de las incógnitas restantes $x_{j_{m+1}}, \ldots, x_{j_n}$. Es decir, que existen $h_{i_1}, \ldots, h_{i_m} : \mathbb{R}^{n-m} \to \mathbb{R}$ tales que

$$x_{i_k} = h_{i_k}(x_{j_{m+1}}, \ldots, x_{j_n})$$

para $k \in \{1, \ldots, m\}$, y que las soluciones del sistema de ecuaciones 5.9 están dadas por las $n-$adas formadas con los n números reales

$$h_{i_1}(x_{j_{m+1}}, \ldots, x_{j_n}), \ldots, h_{i_m}(x_{j_{m+1}}, \ldots, x_{j_n}), x_{j_{m+1}}, \ldots, x_{j_n},$$

en donde $h_{i_k}(x_{j_{m+1}}, \ldots, x_{j_n})$ es la i_k-ésima coordenada y x_{j_l} es la j_l-ésima coordenada, para $k \in \{1, \ldots, m\}$ y $l \in \{m+1, \ldots, n\}$.

5.4.2. El caso no lineal

Una vez que hemos analizado el caso de un sistema de ecuaciones lineales, el siguiente paso será considerar un sistema de ecuaciones, no necesariamente lineales, determinado por m funciones $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ dado por

$$\begin{aligned} g_1(x_1, \ldots, x_n) &= 0 \\ g_2(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ g_m(x_1, \ldots, x_n) &= 0, \end{aligned} \tag{5.13}$$

en donde supondremos que cada función g_i es de clase C^1 en $\mathbb{R}^n$.

Denotaremos por S al conjunto de soluciones de este sistema, es decir

$$S = \{\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid g_1(\hat{x}) = 0, \ldots, g_m(\hat{x}) = 0\}$$

y nuestro objetivo no será encontrar y caracterizar a todos los elemento de S (como hicimos en el caso del sistema de ecuaciones lineales), lo que sin duda es un problema bastante difícil. Nuestro objetivo es algo más "modesto" y consiste en lo siguiente: si tenemos una solución $\hat{x}_0$ del sistema 5.13, es decir que $\hat{x}_0 \in S$, ¿es posible "decir algo" de las soluciones de 5.13 que están "cerca" de $\hat{x}_0$? La respuesta a esta pregunta es justo el teorema de la función

implícita, el cual nos dice "algo" sobre el comportamiento de las soluciones del sistema de ecuaciones 5.13 "alrededor" del punto $\hat{x}_0$.

La idea principal detrás del teorema de la función implícita es la siguiente: dado que nuestro objetivo es conocer el comportamiento de las soluciones del sistema de ecuaciones 5.13 "cerca" o "alrededor" del punto $\hat{x}_0$, entonces sustituyamos cada ecuación $g_i(x_1, \ldots, x_n) = 0$ por su "mejor aproximación lineal en $\hat{x}_0$", es decir por la ecuación

$$Dg_i(\hat{x}_0)(\hat{x} - \hat{x}_0) + g_i(\hat{x}_0) = Dg_i(\hat{x}_0)(\hat{x} - \hat{x}_0) = 0,$$

o lo que es lo mismo, por la ecuación

$$\frac{\partial g_i}{\partial x_1}(\hat{x}_0)x_1 + \frac{\partial g_i}{\partial x_2}(\hat{x}_0)x_2 + \cdots + \frac{\partial g_i}{\partial x_n}(\hat{x}_0)x_n + b_i = 0,$$

en donde $b_i = -Dg_i(\hat{x}_0)(\hat{x}_0)$, para cada $i \in \{1, \ldots, n\}$.

De esta forma, si para el sistema de ecuaciones lineales dado por

$$\begin{aligned}
\frac{\partial g_1}{\partial x_1}(\hat{x}_0)x_1 + \frac{\partial g_1}{\partial x_2}(\hat{x}_0)x_2 + \cdots + \frac{\partial g_1}{\partial x_n}(\hat{x}_0)x_n + b_1 &= 0 \\
\frac{\partial g_2}{\partial x_1}(\hat{x}_0)x_1 + \frac{\partial g_2}{\partial x_2}(\hat{x}_0)x_2 + \cdots + \frac{\partial g_2}{\partial x_n}(\hat{x}_0)x_n + b_2 &= 0 \\
&\vdots \\
\frac{\partial g_m}{\partial x_1}(\hat{x}_0)x_1 + \frac{\partial g_m}{\partial x_2}(\hat{x}_0)x_2 + \cdots + \frac{\partial g_m}{\partial x_n}(\hat{x}_0)x_n + b_m &= 0
\end{aligned} \tag{5.14}$$

se tiene que la matriz

$$\begin{bmatrix}
\frac{\partial g_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0) \\
\vdots & \ddots & \vdots \\
\frac{\partial g_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0)
\end{bmatrix} \tag{5.15}$$

es invertible, por lo visto para el caso de los sistemas de ecuaciones lineales, sabemos que las incógnitas $x_1, \ldots, x_m$ se pueden poner en función de las incógnitas $x_{m+1}, \ldots, x_n$. Es decir que existe $h : \mathbb{R}^{n-m} \to \mathbb{R}^m$ tal

que $(x_1, \ldots, x_m) = h(x_{m+1}, \ldots, x_n)$, y que sus soluciones están dadas por $(h(x_{m+1}, \ldots, x_n), x_{m+1}, \ldots, x_n)$ para todo $(x_{m+1}, \ldots, x_n) \in \mathbb{R}^{n-m}$.

Pues bien, como el sistema de ecuaciones dado por 5.14 se "parece" mucho al sistema de ecuaciones dado por 5.13 "alrededor" o "cerca" de $\hat{x}_0$, el teorema de la función implícita asegura que, "alrededor" o "cerca" de $\hat{x}_0$, las incógnitas $x_1, \ldots, x_m$ también se pueden poner en función de las incógnitas $x_{m+1}, \ldots, x_n$; es decir, si

$$\hat{x}_0 = \left(x_1^{(0)}, \ldots, x_m^{(0)}, x_{m+1}^{(0)}, \ldots, x_n^{(0)}\right),$$

este teorema nos asegura que existen $\delta > 0$, $V \subset \mathbb{R}^{n-m}$ un conjunto abierto, y una función (que resultará ser única)

$$h : V \subset \mathbb{R}^{n-m} \to \mathbb{R}^m$$

de clase C^1, tales que $\left(x_{m+1}^{(0)}, \ldots, x_n^{(0)}\right) \in V$,

$$h\left(x_{m+1}^{(0)}, \ldots, x_n^{(0)}\right) = \left(x_1^{(0)}, \ldots, x_m^{(0)}\right)$$

y las $n-$adas $(h(x_{m+1}, \ldots, x_n), x_{m+1}, \ldots, x_n)$ también son soluciones del sistema 5.13. Es decir, que

$$(h(x_{m+1}, \ldots, x_n), x_{m+1}, \ldots, x_n) \in B_\delta(\hat{x}_0) \cap S$$

para cada $(x_{m+1}, \ldots, x_n) \in V$.

Antes de dar la formulación más precisa del teorema de la función implícita, haremos unas observaciones importantes. La primera de ellas tiene que ver con la elección de la matriz dada por 5.15. Como en el caso del sistema de ecuaciones lineales, si en general los índices $1 \leq i_1 < i_2 < \cdots < i_m \leq n$, $1 \leq j_{m+1} < j_{m+2} < \cdots < j_n \leq n$, con $\{i_1, \ldots, i_m\} \cap \{j_{m+1}, \ldots, j_n\} = \varnothing$, son tales que la matriz

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_{i_1}}(\hat{x}_0) & \cdots & \frac{\partial g_1}{\partial x_{i_m}}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_{i_1}}(\hat{x}_0) & \cdots & \frac{\partial g_m}{\partial x_{i_m}}(\hat{x}_0) \end{bmatrix}$$

es invertible, entonces lo que el teorema de la función implícita afirma es que "alrededor" de $\hat{x}_0 \in S$ las incógnitas $x_{i_1}, \ldots, x_{i_m}$ se pueden poner en función de las incógnitas restantes $x_{j_{m+1}}, \ldots, x_{j_n}$, es decir que existen $\delta > 0$, $V \subset \mathbb{R}^{n-m}$ un conjunto abierto, y funciones

$$h_{i_1}, \ldots, h_{i_m} : V \subset \mathbb{R}^{n-m} \to \mathbb{R}$$

(que serán únicas) tales que $\left(x^{(0)}_{j_{m+1}}, \ldots, x^{(0)}_{j_n}\right) \in V$,

$$x^{(0)}_{i_k} = h_{i_k}\left(x^{(0)}_{j_{m+1}}, \ldots, x^{(0)}_{j_n}\right)$$

para $k \in \{1, \ldots, m\}$, y que las $n-$adas formadas con los n números reales

$$h_{i_1}(x_{j_{m+1}}, \ldots, x_{j_n}), \ldots, h_{i_m}(x_{j_{m+1}}, \ldots, x_{j_n}), x_{j_{m+1}}, \ldots, x_{j_n},$$

en donde $h_{i_k}(x_{j_{m+1}}, \ldots, x_{j_n})$ es la i_k-ésima coordenada y x_{j_l} es la j_l-ésima coordenada, para $k \in \{1, \ldots, m\}$ y $l \in \{m+1, \ldots, n\}$, pertenecen al conjunto $B_\delta(\hat{x}_0) \cap S$ para cada $\left(x_{j_{m+1}}, \ldots, x_{j_n}\right) \in V$.

Las otras observaciones que haremos en realidad son interpretaciones geométricas del teorema de la función implícita para los casos de $\mathbb{R}^2$ y $\mathbb{R}^3$. El caso más sencillo es el de $\mathbb{R}^2$, en el que sólo se puede tener una restricción de la forma

$$g(x, y) = 0,$$

que no es más que el conjunto de nivel 0 de g ($N_0(g)$).

De acuerdo con lo visto anteriormente, si $\hat{x}_0 = (x_0, y_0) \in \mathbb{R}^2$ es tal que $g(x_0, y_0) = 0$ y $\frac{\partial g}{\partial x}(\hat{x}_0) \neq 0$ (es decir, que la matriz de 1×1 dada por $\left[\frac{\partial g}{\partial x}(\hat{x}_0)\right]$ ¡es invertible!), entonces para puntos $(x, y) \in N_0(g)$ "cercanos" al punto $\hat{x}_0 = (x_0, y_0)$, su coordenada (o variable) x se puede poner en función de su coordenada (o variable) y, es decir que existen $\delta > 0$, $I \subset \mathbb{R}$ abierto, y $h : I \subset \mathbb{R} \to \mathbb{R}$ tales que $y_0 \in I$, $h(y_0) = x_0$ y $(h(y), y) \in B_\delta(\hat{x}_0)) \cap N_0(g)$ para toda $y \in I$. Análogamente, si $\frac{\partial g}{\partial y}(\hat{x}_0) \neq 0$, entonces para puntos de $N_0(g)$ "cercanos" al punto $\hat{x}_0 = (x_0, y_0)$ su coordenada (o variable) y se puede poner en función de la coordenada (o variable) x, es decir que existen

$\delta > 0$, $I \subset \mathbb{R}$ abierto, y $h : I \subset \mathbb{R} \to \mathbb{R}$ tales que $x_0 \in I$, $h(x_0) = y_0$ y $(x, h(x)) \in B_\delta(\hat{x}_0) \cap N_0(g)$ para toda $x \in I$.

En términos geométricos, lo anterior significa que si el vector $\nabla g(\hat{x}_0)$ no es horizontal ($\frac{\partial g}{\partial y}(\hat{x}_0) \neq 0$), es decir no es paralelo al eje X, entonces un "sector" de la curva de nivel 0 de g "alrededor" del punto $\hat{x}_0 = (x_0, y_0)$ se puede "ver" como la gráfica de una función de la forma $y = h(x)$ (ver figura 5.4 (a)). Y si el vector $\nabla g(\hat{x}_0)$ no es vertical ($\frac{\partial g}{\partial x}(\hat{x}_0) \neq 0$), es decir no es paralelo al eje Y, entonces un sector de la curva de nivel 0 de g alrededor del punto $\hat{x}_0 = (x_0, y_0)$ se puede ver como la gráfica de una función de la forma $x = h(y)$ (ver figura 5.4 (b)).

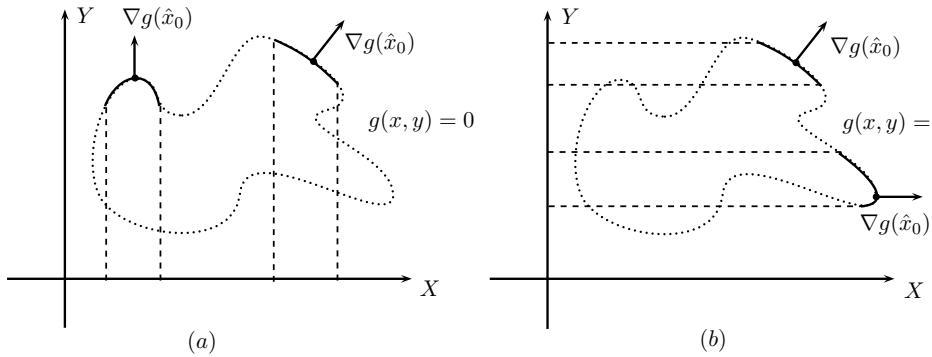


Figura 5.4: Si el vector $\nabla g(\hat{x}_0)$ no es horizontal (como en la figura (a)), entonces un sector de la curva de nivel 0 de g alrededor del punto $\hat{x}_0$ se puede ver como la gráfica de una función de la forma $x = h(y)$. Y si el vector $\nabla g(\hat{x}_0)$ no es vertical (como en la figura (b)), entonces un sector de la curva de nivel alrededor del punto $\hat{x}_0$ se puede ver como la gráfica de una función de la forma $y = h(x)$.

Para el caso de $\mathbb{R}^3$, si sólo se tiene una "restricción" de la forma

$$g(x, y, z) = 0$$

y $\hat{x}_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$ es tal que $g(x_0, y_0, z_0) = 0$ y $\frac{\partial g}{\partial x}(\hat{x}_0) \neq 0$ (lo que significa que el vector $\nabla g(\hat{x}_0)$ no está en el plano YZ), entonces para puntos

$(x, y, z) \in N_0(g)$ "cercanos" al punto $\hat{x}_0$, la coordenada (o variable) x se puede poner en función de las correspondientes coordenadas (o variables) y y z, es decir que existen $\delta > 0$, $V \subset \mathbb{R}^2$ abierto, y $h : V \subset \mathbb{R}^2 \to \mathbb{R}$ tales que $(y_0, z_0) \in V$, $h(y_0, z_0) = x_0$ y $(h(y, z), y, z) \in B_\delta(\hat{x}_0) \cap N_0(g)$ para toda $(y, z) \in V$.

Lo anterior, nuevamente en términos geométricos, significa que un "pedazo" del conjunto de nivel 0 de g que contiene al punto $\hat{x}_0 = (x_0, y_0, z_0)$ se puede "ver" como la gráfica de una función de la forma $x = h(y, z)$ (ver figura 5.5). Los otros casos ($\frac{\partial g}{\partial y}(\hat{x}_0) \neq 0$ y $\frac{\partial g}{\partial z}(\hat{x}_0) \neq 0$) tienen interpretaciones geométricas semejantes.

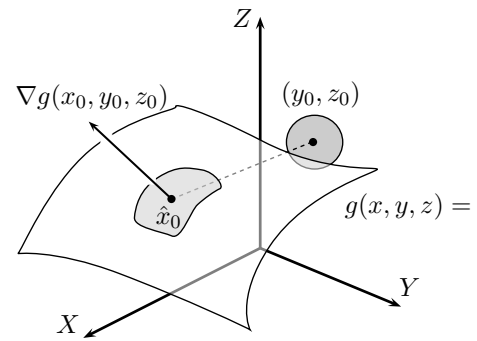


Figura 5.5: Si $\frac{\partial g}{\partial x}(\hat{x}_0) \neq 0$ (lo que significa que el vector $\nabla g(\hat{x}_0)$ no está en el plano YZ), entonces un sector de la superficie de nivel 0 de g alrededor del punto $\hat{x}_0 = (x_0, y_0, z_0)$ se puede ver como la gráfica de una función de la forma $x = h(y, z)$.

La otra posibilidad en $\mathbb{R}^3$ es cuando tenemos dos restricciones de la forma

$$g_1(x, y, z) = 0$$
$$g_2(x, y, z) = 0,$$

en cuyo caso el conjunto S de soluciones de este sistema se ve como una curva.

Si $\hat{x}_0 = (x_0, y_0, z_0) \in S$, nótese que el sistema de ecuaciones dado por 5.14 se reduce al sistema de dos ecuaciones

$$\frac{\partial g_1}{\partial x}(\hat{x}_0)x + \frac{\partial g_1}{\partial y}(\hat{x}_0)y + \frac{\partial g_1}{\partial y}(\hat{x}_0)z - \nabla g_1(\hat{x}_0)\cdot\hat{x}_0 = 0$$
$$\frac{\partial g_2}{\partial x}(\hat{x}_0)x + \frac{\partial g_2}{\partial y}(\hat{x}_0)y + \frac{\partial g_2}{\partial y}(\hat{x}_0)z - \nabla g_2(\hat{x}_0)\cdot\hat{x}_0 = 0$$

y el determinante de cada una de las tres posibles submatrices de 2×2 que se pueden construir a partir de este sistema coinciden (salvo posiblemente por el signo), con las coordenadas del vector

$$\begin{aligned}\nabla g_1(\hat{x}_0)\times\nabla g_2(\hat{x}_0) = &\det\begin{bmatrix}\frac{\partial g_1}{\partial y}(\hat{x}_0) & \frac{\partial g_1}{\partial z}(\hat{x}_0)\\ \frac{\partial g_2}{\partial y}(\hat{x}_0) & \frac{\partial g_2}{\partial z}(\hat{x}_0)\end{bmatrix}\hat{e}_1\\ &-\det\begin{bmatrix}\frac{\partial g_1}{\partial x}(\hat{x}_0) & \frac{\partial g_1}{\partial z}(\hat{x}_0)\\ \frac{\partial g_2}{\partial x}(\hat{x}_0) & \frac{\partial g_2}{\partial z}(\hat{x}_0)\end{bmatrix}\hat{e}_2\\ &+\det\begin{bmatrix}\frac{\partial g_1}{\partial x}(\hat{x}_0) & \frac{\partial g_1}{\partial y}(\hat{x}_0)\\ \frac{\partial g_2}{\partial x}(\hat{x}_0) & \frac{\partial g_2}{\partial y}(\hat{x}_0)\end{bmatrix}\hat{e}_3\end{aligned}$$

De esta forma, si la matriz

$$\begin{bmatrix}\frac{\partial g_1}{\partial y}(\hat{x}_0) & \frac{\partial g_1}{\partial z}(\hat{x}_0)\\ \frac{\partial g_2}{\partial y}(\hat{x}_0) & \frac{\partial g_2}{\partial z}(\hat{x}_0)\end{bmatrix}$$

tiene determinante distinto de 0 (es decir, si la primera coordenada del vector $\nabla g_1(\hat{x}_0)\times\nabla g_2(\hat{x}_0)$ es distinta de 0, o equivalentemente, que este vector no pertenece al plano YZ), entonces las coordenadas (o variables) y y z de puntos de S cercanos al punto $\hat{x}_0$ se pueden poner en función de la coordenada (o variable) x. Es decir, existen $\delta>0$, $I\subset\mathbb{R}$ abierto, y $h_1, h_2 : I\subset\mathbb{R}\to\mathbb{R}$ tales que $x_0\in I$, $h_1(x_0)=y_0$, $h_2(x_0)=z_0$ y $(x,h_1(x),h_2(x))\in B_\delta(\hat{x}_0)\cap S$ para toda $x\in I$. Lo anterior significa que la función de $\mathbb{R}$ en $\mathbb{R}^3$ dada por $h(x)=(x,h_1(x),h_2(x))$ (para $x\in I$) es una parametrización de un "pedazo" de la curva determinada por las restricciones g_1 y g_2 (ver figura 5.6). Las otras dos posibilidades, correspondientes a las otras dos matrices de 2×2 que se pueden obtener con las derivadas parciales de g_1 y g_2, se interpretan de manera análoga.

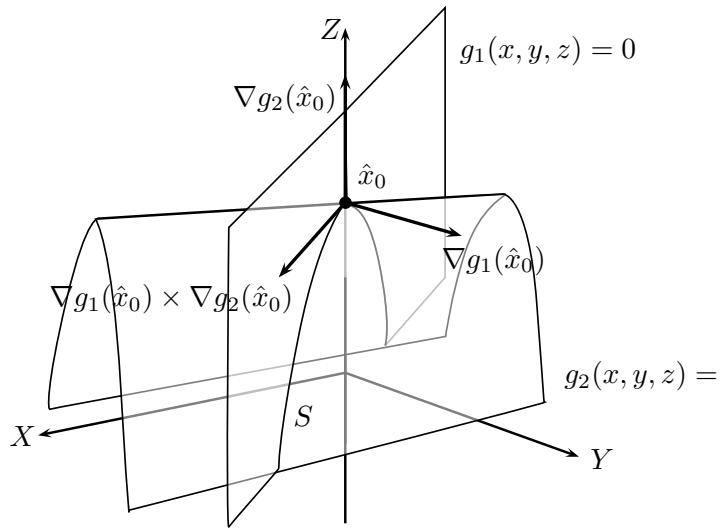


Figura 5.6: Si la primera coordenada del vector $\nabla g_1(\hat{x}_0) \times \nabla g_2(\hat{x}_0)$ es distinta de 0, entonces las variables y y z alrededor del punto $\hat{x}_0 = (x_0, y_0, z_0)$ se pueden poner en función de la variable x. Esto significa que un sector de la curva S (determinada por la intersección de las restricciones g_1 y g_2) se puede parametrizar en términos de la variable x.

Las dos primeras interpretaciones geométricas que acabamos de hacer, los casos en que el conjunto S coincide con un conjunto de nivel en $\mathbb{R}^2$ o uno en $\mathbb{R}^3$, son justo los hechos geométricos que se mencionaron en los comentarios que hicimos posteriores al ejemplo 4.10 del capítulo 4. En ese ejemplo calculamos la recta tangente de un cierto conjunto de nivel en $\mathbb{R}^2$, y el plano tangente de un cierto conjunto de nivel en $\mathbb{R}^3$, de acuerdo con las definiciones 4.7 y 4.4. Y precisamente, adelantándonos a las interpretaciones geométricas del teorema de la función implícita que acabamos de hacer, comentamos que esa recta y ese plano también se podían obtener como la recta tangente a la gráfica de una función de $\mathbb{R}$ en $\mathbb{R}$, y el plano tangente a la gráfica de una función de $\mathbb{R}^2$ en $\mathbb{R}$, respectivamente. Lo que ahora nos proponemos es tomar esos mismos conjuntos y mostrar que estas afirmaciones son ciertas.

Ejemplo 5.3

1. *Consideremos el conjunto de nivel* 0 *(*$N_0(g)$*) de la función de clase* C^1 *en* $\mathbb{R}^2$

$$g(x,y) = (x^2+y^2-2x)^2 - 4(x^2+y^2)$$

que, como se mencionó en el inciso 1 del ejemplo 4.10 del capítulo 4, es la curva cardioide.

Como

$$\frac{\partial g}{\partial x}(x,y) = 2(x^2+y^2-2x)\,(2x-2) - 8x$$

y

$$\frac{\partial g}{\partial y}(x,y) = 2(x^2+y^2-2x)2y - 8y$$

para cada $(x,y) \in \mathbb{R}^2$*, en particular para el punto* $(0,2) \in N_0(g)$ *se tiene que* $\frac{\partial g}{\partial x}(0,2) = -16 \neq 0$ *y* $\frac{\partial g}{\partial y}(0,2) = 16 \neq 0$*.*

Dado que ambas derivadas parciales son distintas de 0*, podemos aplicar el teorema de la función implícita para los dos casos.*

Si consideramos el hecho de que $\frac{\partial g}{\partial x}(0,2) \neq 0$*, el teorema nos asegura que existen* $\delta > 0$*,* $I \subset \mathbb{R}$ *abierto, y* $h : I \subset \mathbb{R} \to \mathbb{R}$ *de clase* C^1 *tales que* $2 \in I$*,* $h(2) = 0$ *y* $(h(t),t)) \in B_\delta((0,2)) \cap N_0(g)$ *para toda* $t \in I$*.*

En particular se tiene que

$$g\left(h\left(t\right),t\right) = 0 \tag{5.16}$$

para toda $t \in I$*. Derivando esta identidad usando la regla de la cadena obtenemos que*

$$\frac{\partial g}{\partial x}\left(h\left(t\right),t\right) h'\left(t\right) + \frac{\partial g}{\partial y}\left(h\left(t\right),t\right) = 0$$

para toda $t \in I$*.*

Así, evaluando para $t = 2$, se tiene que

$$\begin{aligned}\frac{\partial g}{\partial x}\left(h\left(2\right),2\right)h'\left(2\right) &= \frac{\partial g}{\partial x}\left(0,2\right)h'\left(2\right)\\ &= -\frac{\partial g}{\partial y}\left(0,2\right)\end{aligned}$$

y por lo tanto, como $\frac{\partial g}{\partial x}(0,2) \neq 0$, concluimos que

$$\begin{aligned}h'\left(2\right) &= -\frac{\frac{\partial g}{\partial y}\left(0,2\right)}{\frac{\partial g}{\partial x}\left(0,2\right)}\\ &= -\frac{16}{-16}\\ &= 1.\end{aligned}$$

Una vez que llegamos a este punto, es importante hacer la siguiente observación: en estricto sentido, la gráfica de la función h es el conjunto de parejas de la forma $(t, h(t)) \in \mathbb{R}^2$, que no son *las parejas que pertenecen al conjunto de nivel 0 de la función g.*

De acuerdo con la identidad 5.16, hay que permutar las coordenadas de los elementos de la gráfica de h para que dichas parejas pertenezcan al conjunto $N_0(g)$.

La explicación de este hecho es que, en este caso, es la coordenada x a la que pusimos en función de la coordenada y. Para ser congruentes con lo anterior y no caer en errores, al momento de escribir la ecuación de la recta tangente a la gráfica de h en el punto $(0,2)$, habrá que hacer lo mismo, escribir a la coordenada x en función de la coordenada y. Es decir, escribir que

$$\begin{aligned}x &= h'\left(2\right)\left(y-2\right)+h\left(2\right)\\ &= y-2+0,\end{aligned}$$

o equivalentemente

$$y = x + 2.$$

Procediendo de esta forma, sí obtenemos la misma recta que en el inciso 1 del ejemplo 4.10 del capítulo 4.

2. *Consideremos el conjunto de nivel* 0 *($N_0(g)$) de la función de clase* C^1 *en* $\mathbb{R}^3$

$$g(x,y,z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1$$

que, como se mencionó en el inciso 2 del ejemplo 4.10 del capítulo 4, es un elipsoide.

En este caso tenemos que

$$\frac{\partial g}{\partial x}(x,y,z) = 2\frac{x}{a^2}$$
$$\frac{\partial g}{\partial y}(x,y,z) = 2\frac{y}{b^2}$$
$$\frac{\partial g}{\partial z}(x,y,z) = 2\frac{z}{c^2},$$

de tal forma que si elegimos el punto $(0,b,0) \in N_0(g)$*, entonces*

$$\frac{\partial g}{\partial x}(0,b,0) = 0$$
$$\frac{\partial g}{\partial y}(0,b,0) = \frac{2}{b}$$
$$\frac{\partial g}{\partial z}(0,b,0) = 0.$$

Por tanto, como en este caso la $\frac{\partial g}{\partial y}(0,b,0) = \frac{2}{b}$ *es la única derivada parcial que es distinta de* 0 *en el punto* $(0,b,0)$*, el teorema de la función implícita sólo nos permite asegurar que para puntos* $(x,y,z) \in N_0(g)$ *que están "cercanos" al punto* $(0,b,0)$*, su coordenada* y *se puede poner en función de sus coordenadas* x *y* z*. Es decir, que existen* $\delta > 0$*,* $V \subset \mathbb{R}^2$*, y* $h : V \subset \mathbb{R}^2 \to \mathbb{R}$ *tales que* $(0,0) \in V$*,* $h(0,0) = b$ *y* $(x, h(x,z), z) \in B_\delta((0,b,0)) \cap N_0(g)$ *para toda* $(x,z) \in V$*.*

Nótese nuevamente, como en el inciso anterior, que en estricto sentido las ternas que pertenecen a la gráfica de la función h son las de la forma $(x, z, h(x, z))$. Las ternas $(x, h(x, z), z)$ se obtienen al hacer una permutación de las coordenadas de las ternas $(x, z, h(x, z))$ (intercambiando la segunda coordenada con la tercera), y son las que pertenecen al conjunto $N_0(g)$.

Dado que para calcular el plano tangente a la gráfica de la función h en el punto $(0, b, 0)$ es necesario calcular sus derivadas parciales en el punto $(0, 0)$, hacemos esto derivando nuevamente la identidad anterior por medio de la regla de la cadena.

En efecto, como el lado izquierdo de la identidad anterior es la composición de la función $H : V \subset \mathbb{R}^2 \to \mathbb{R}^3$ dada por

$$H(x, z) = (x, h(x, z), z)$$

seguida de la función g, y esta composición nos da la función constante 0, aplicando la regla de la cadena obtenemos que

$$\begin{aligned}
&D(g \circ H)(x, z) \\
&= Dg(H(x, z))\, DH(x, z) \\
&= \left[\frac{\partial g}{\partial x}(H(x, z)) \ \frac{\partial g}{\partial y}(H(x, z)) \ \frac{\partial g}{\partial z}(H(x, z)) \right] \begin{bmatrix} 1 & 0 \\ \frac{\partial h}{\partial x}(x, z) & \frac{\partial h}{\partial z}(x, z) \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 \end{bmatrix}
\end{aligned}$$

para toda $(x, z) \in V$.
Por lo tanto, evaluando esta identidad de matrices en el punto $(0, 0)$ se obtiene que

$$0 = \frac{\partial g}{\partial x}(0, b, 0) + \frac{\partial g}{\partial y}(0, b, 0) \frac{\partial h}{\partial x}(0, 0) = \frac{2}{b} \frac{\partial h}{\partial x}(0, 0),$$

y

$$0 = \frac{\partial g}{\partial y}(0, b, 0) \frac{\partial h}{\partial z}(0, 0) + \frac{\partial g}{\partial z}(0, b, 0) = \frac{2}{b} \frac{\partial h}{\partial z}(0, 0)$$

de donde

$$\frac{\partial h}{\partial x}(0,0) = 0 = \frac{\partial h}{\partial z}(0,0).$$

Si ahora calculamos el plano tangente a la gráfica de h en el punto $(0,0)$, de acuerdo con la definición 4.4 del capítulo 4 y, como en el inciso anterior, recordando que la coordenada y es la que se debe escribir en términos de las coordenadas x y z, se tiene que

$$\begin{aligned} y &= \frac{\partial h}{\partial x}(0,0)(x-0) + \frac{\partial h}{\partial z}(0,0)(z-0) + h(0,0) \\ &= b. \end{aligned}$$

Como el lector podrá comprobar, esta ecuación es la misma que se obtiene en el inciso 2. del ejemplo 4.10 del capítulo 4, tomando $(x_0, y_0, z_0) = (0, b, 0)$.

Una vez dicho y hecho todo lo anterior, escribiremos el tan mencionado teorema de la función implícita. Con el fin de hacer sencilla su redacción, dado que en la primera observación que hicimos ya mencionamos cuál sería su formulación más general, supondremos que la matriz de $m \times m$ que se necesita que sea invertible (como parte de las hipótesis), es la correspondiente a las primeras m variables (la matriz dada en 5.15). Para simplificar an más esta redacción, denotaremos al espacio $\mathbb{R}^n$ como $\mathbb{R}^m \times \mathbb{R}^k$, y si escribimos que $(\hat{x}, \hat{y}) \in \mathbb{R}^m \times \mathbb{R}^k$, esto significa que esta pareja es de la forma

$$(\hat{x}, \hat{y}) = (x_1, \ldots, x_m, y_1, \ldots, y_k).$$

Teorema 5.1 (de la función implícita) *Sean $g_1, \ldots, g_m : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ de clase C^1 en U. Si*

$$S = \left\{ (\hat{x}, \hat{y}) \in U \subset \mathbb{R}^m \times \mathbb{R}^k \mid g_1(\hat{x}, \hat{y}) = 0, \ldots, g_m(\hat{x}, \hat{y}) = 0 \right\}$$

y

$$(\hat{x}_0, \hat{y}_0) = \left(x_1^{(0)}, \ldots, x_m^{(0)}, y_1^{(0)}, \ldots, y_k^{(0)} \right) \in S$$

es tal que la matriz

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

es invertible, entonces existen $\delta > 0$, $V \subset \mathbb{R}^k$ un conjunto abierto, y una función $h : V \subset \mathbb{R}^k \to \mathbb{R}^m$ de clase C^1 en V, tales que $\hat{y}_0 \in V$, $h(\hat{y}_0) = \hat{x}_0$ y $(h(\hat{y}), \hat{y}) \in B_\delta((\hat{x}_0, \hat{y}_0)) \cap S$ para toda $\hat{y} \in V$. Estas propiedades de h la hacen única.

Por razones de comodidad, dejaremos pendiente la prueba de este teorema hasta la siguiente sección, en la que probaremos el teorema de la función inversa y con base en el cual demostraremos el teorema anterior. A cambio de esta prueba, apoyados en el teorema de la función implícita haremos una prueba que en el capítulo 4 dejamos pendiente: la prueba del teorema de los multiplicadores de Lagrange.

La formulación que daremos a continuación de este teorema, aunque totalmente equivalente a la que dimos en el capítulo 4, será escrita de tal forma que esté más acorde con la formulación que acabamos de dar del teorema de la función implícita.

Teorema 5.2 (de los multiplicadores de Lagrange) *Sean:*

1. $g_1, \ldots, g_m : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ *funciones de clase C^1 en U.*
2. $S \subset \mathbb{R}^m \times \mathbb{R}^k$ *el conjunto dado por*

$$S = \left\{ (\hat{x}, \hat{y}) \in U \subset \mathbb{R}^m \times \mathbb{R}^k \mid g_1(\hat{x}, \hat{y}) = 0, \ldots, g_m(\hat{x}, \hat{y}) = 0 \right\}$$

3. $(\hat{x}_0, \hat{y}_0) \in S$ *tal que $\nabla g_1(\hat{x}_0, \hat{y}_0), \ldots, \nabla g_m(\hat{x}_0, \hat{y}_0)$ son linealmente independientes.*

   Si $f : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ es una función de clase C^1 en U tal que f tiene un máximo o mínimo (local) en $(\hat{x}_0, \hat{y}_0)$ sobre S, entonces existen $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ tales que

$$\nabla f(\hat{x}_0, \hat{y}_0) = \lambda_1 \nabla g_1(\hat{x}_0, \hat{y}_0) + \cdots + \lambda_m \nabla g_m(\hat{x}_0, \hat{y}_0).$$

Demostración. Dado que los vectores $\nabla g_1(\hat{x}_0, \hat{y}_0), \ldots, \nabla g_m(\hat{x}_0, \hat{y}_0)$ son linealmente independientes, se tiene que el rango por renglones de la matriz

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_n}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_n}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

debe ser m, y como éste debe ser igual a su rango por columnas (inciso (c) del Corolario 2 del Teorema 3.6 de la referencia [2]), dicha matriz debe tener m columnas linealmente independientes.

Supondremos, sin pérdida de generalidad, que estas columnas son las primeras m de tal forma que la matriz de $m \times m$

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

será invertible. De esta forma, por el teorema de la función implícita, sabemos que existen $\delta > 0$, $V \subset \mathbb{R}^k$ y

$$h = (h_1, \ldots, h_m) : V \subset \mathbb{R}^k \to \mathbb{R}^m$$

de clase C^1 en V tales que $\hat{y}_0 \in V$, $h(\hat{y}_0) = \hat{x}_0$ y $(h(\hat{y}), \hat{y}) \in B_\delta((\hat{x}_0, \hat{y}_0)) \cap S$ para toda $\hat{y} \in V$.

Definamos ahora la función

$$H : V \subset \mathbb{R}^k \to \mathbb{R}^m \times \mathbb{R}^k$$

como $H(\hat{y}) = (h(\hat{y}), \hat{y})$, la cual es de clase C^1 en V. Ahora, como $(h(\hat{y}), \hat{y}) \in S$ para toda $\hat{y} \in V$, se tiene que $(g_i \circ H)(\hat{y}) = 0$ para toda $\hat{y} \in V$ de modo que en particular

$$\begin{aligned} D(g_i \circ H)(\hat{y}_0) &= Dg_i(H(\hat{y}_0))\, DH(\hat{y}_0) \\ &= Dg_i(\hat{x}_0, \hat{y}_0)\, DH(\hat{y}_0) \\ &= \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} \in M_{1\times k}(\mathbb{R}). \end{aligned}$$

Es decir, que el producto de las matrices

$$\left[\begin{array}{cccccc} \frac{\partial g_i}{\partial x_1}(\hat{x}_0,\hat{y}_0) & \cdots & \frac{\partial g_i}{\partial x_m}(\hat{x}_0,\hat{y}_0) & \frac{\partial g_i}{\partial y_1}(\hat{x}_0,\hat{y}_0) & \cdots & \frac{\partial g_i}{\partial y_k}(\hat{x}_0,\hat{y}_0) \end{array}\right] \tag{5.17}$$

y

$$\left[\begin{array}{ccc} \frac{\partial h_1}{\partial y_1}(\hat{y}_0) & \cdots & \frac{\partial h_1}{\partial y_k}(\hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial y_1}(\hat{y}_0) & \cdots & \frac{\partial h_m}{\partial y_k}(\hat{y}_0) \\ 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{array}\right] \tag{5.18}$$

es la matriz $\left[\begin{array}{ccc} 0 & \cdots & 0 \end{array}\right] \in M_{1\times k}(\mathbb{R})$, para cada $i \in \{1,\ldots,m\}$.

Si ahora definimos los vectores

$$\hat{w}_j = \left(\left(\frac{\partial h_1}{\partial y_j}(\hat{y}_0),\ldots,\frac{\partial h_m}{\partial y_j}(\hat{y}_0)\right),\hat{e}_j\right) \in \mathbb{R}^m \times \mathbb{R}^k$$

para $j \in \{1,\ldots,k\}$ (en donde $\hat{e}_1,\ldots,\hat{e}_k \in \mathbb{R}^k$ son los vectores canónicos), se tiene que $\hat{w}_1,\ldots,\hat{w}_k$ son linealmente independientes, de modo que si $W \subset \mathbb{R}^m \times \mathbb{R}^k$ es el subespacio generado por estos vectores (es decir, que $W = \langle \hat{w}_1,\ldots,\hat{w}_k\rangle$), entonces W es de dimensión k ($\dim(W) = k$).

Ahora nótese que del hecho de que el producto de las matrices 5.17 y 5.18 sea la matriz (de $1 \times k$) idénticamente cero se concluye que

$$\nabla g_i(\hat{x}_0,\hat{y}_0)\cdot\hat{w}_j = 0$$

para cada $i \in \{1,\ldots,m\}$ y cada $j \in \{1,\ldots,k\}$. De esta forma, se tiene que $\nabla g_i(\hat{x}_0,\hat{y}_0) \in W^{\perp}$ (el complemento ortogonal de W en $\mathbb{R}^m \times \mathbb{R}^k$, de acuerdo con la definición de la página 349 de la referencia [2]) para cada $i \in \{1,\ldots,m\}$. Por otra parte, por el inciso (c) del Teorema 6.7 de la referencia [2], se tiene que

$$\begin{aligned}\dim(W^{\perp}) &= \dim\left(\mathbb{R}^m \times \mathbb{R}^k\right) - \dim(W)\\ &= m + k - k\\ &= m\end{aligned}$$

y como $\nabla g_1(\hat{x}_0, \hat{y}_0), \ldots, \nabla g_m(\hat{x}_0, \hat{y}_0)$ son linealmente independientes, concluimos que estos vectores son una base de $W^{\perp}$.

Finalmente, si ahora consideramos la función $f \circ H$ (la cual está definida en alguna vecindad del punto $\hat{y}_0 \in \mathbb{R}^k$), dado que f tiene un valor extremo (local) en el punto $(\hat{x}_0, \hat{y}_0) = H(\hat{y}_0)$, entonces $f \circ H$ tiene un valor extremo (local) en el punto $\hat{y}_0$ (problema 43 del capítulo 4), de modo que $\hat{y}_0$ es un punto crítico de $f \circ H$ y por lo tanto

$$\begin{aligned}D(f \circ H)(\hat{y}_0) &= Df\left(H(\hat{y}_0)\right) DH(\hat{y}_0)\\ &= Df\left(\hat{x}_0, \hat{y}_0\right) DH(\hat{y}_0)\\ &= \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} \in M_{1\times k}(\mathbb{R}).\end{aligned}$$

Es decir, que el producto de las matrices

$$\begin{bmatrix} \frac{\partial f}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial f}{\partial x_m}(\hat{x}_0, \hat{y}_0) & \frac{\partial f}{\partial y_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial f}{\partial y_k}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

y

$$\begin{bmatrix} \frac{\partial h_1}{\partial y_1}(\hat{y}_0) & \cdots & \frac{\partial h_1}{\partial y_k}(\hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial y_1}(\hat{y}_0) & \cdots & \frac{\partial h_m}{\partial y_k}(\hat{y}_0) \\ 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

es la matriz $\begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} \in M_{1\times k}(\mathbb{R})$.

Por tanto, también se tiene que

$$\nabla f\left(\hat{x}_0, \hat{y}_0\right) \cdot \hat{w}_j = 0$$

para cada $j \in \{1, \ldots, k\}$, de modo que $\nabla f(\hat{x}_0, \hat{y}_0)$ también pertenece a $W^\perp = \langle \hat{w}_1, \ldots, \hat{w}_k \rangle^\perp$, y como $\nabla g_1(\hat{x}_0, \hat{y}_0), \ldots, \nabla g_m(\hat{x}_0, \hat{y}_0)$ son una base de $W^\perp$, entonces $\nabla f(\hat{x}_0, \hat{y}_0)$ debe ser una combinación lineal de estos vectores, es decir, existen $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ tales que

$$\nabla f(\hat{x}_0, \hat{y}_0) = \lambda_1 \nabla g_1(\hat{x}_0, \hat{y}_0) + \cdots + \lambda_m \nabla g_m(\hat{x}_0, \hat{y}_0),$$

con lo cual concluimos la prueba. ■

5.5. El teorema de la función inversa

Además del papel importante que juega en la prueba del teorema de la función implícita, el teorema de la función inversa es relevante por méritos propios. De hecho, este teorema no debe ser ajeno al lector, pues el caso particular de este teorema para funciones de $\mathbb{R}$ en $\mathbb{R}$ forma parte de los resultados importantes de un primer curso de cálculo.

En esta sección haremos uso de todo el material desarrollado en las primeras secciones de este capítulo, pero adicionalmente necesitaremos algunos resultados relacionados con las funciones lineales de $\mathbb{R}^n$ en $\mathbb{R}^m$ y su representación matricial, los cuales introduciremos a continuación.

Si $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ es una función de clase C^1 en su dominio U, sabemos que para cada $\hat{x} \in U$ existe la derivada de f, que la derivada es una función lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$, que ésta se representa por una matriz (la cual depende de las bases que se elijan para $\mathbb{R}^n$ en $\mathbb{R}^m$, y a la que llamamos matriz jacobiana), y que para cada $\hat{x} \in U$, esta matriz está dada por

$$Df(\hat{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\hat{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\hat{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\hat{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\hat{x}) \end{bmatrix}.$$

Como se podrá observar (y era de esperarse), para cada $\hat{x} \in U$ tenemos asociada una matriz cuyas entradas son funciones continuas de $\hat{x}$, y una cuestión importante es determinar de qué forma se refleja este hecho en las

funciones lineales que representan estas matrices. Para ello, observemos que si evaluamos dos de estas funciones lineales (es decir, la derivada de f en dos puntos $\hat{x}, \hat{x}_0 \in U$) en un punto arbitrario $\hat{z} = (z_1, \ldots, z_n) \in \mathbb{R}^n$, se tiene que la diferencia entre estos valores está dada por

$$\begin{aligned}
&Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z}) \\
&= (Df(\hat{x}) - Df(\hat{x}_0))\,(\hat{z}) \\
&= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\hat{x}) - \frac{\partial f_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\hat{x}) - \frac{\partial f_1}{\partial x_n}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\hat{x}) - \frac{\partial f_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\hat{x}) - \frac{\partial f_m}{\partial x_n}(\hat{x}_0) \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \\
&= ((\nabla f_1(\hat{x}) - \nabla f_1(\hat{x}_0)) \cdot \hat{z}, \ldots, (\nabla f_m(\hat{x}) - \nabla f_m(\hat{x}_0)) \cdot \hat{z})\,, \qquad (5.19)
\end{aligned}$$

de tal forma que su distancia satisface que

$$\begin{aligned}
&\|Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z})\| \\
&= \|((\nabla f_1(\hat{x}) - \nabla f_1(\hat{x}_0)) \cdot \hat{z}, \ldots, (\nabla f_m(\hat{x}) - \nabla f_m(\hat{x}_0)) \cdot \hat{z})\| \\
&\leq |(\nabla f_1(\hat{x}) - \nabla f_1(\hat{x}_0)) \cdot \hat{z}| + \cdots + |(\nabla f_m(\hat{x}) - \nabla f_m(\hat{x}_0)) \cdot \hat{z}| \\
&\leq \|\nabla f_1(\hat{x}) - \nabla f_1(\hat{x}_0)\|\,\|\hat{z}\| + \cdots + \|\nabla f_m(\hat{x}) - \nabla f_m(\hat{x}_0)\|\,\|\hat{z}\| \\
&= \|\hat{z}\| \sum_{j=1}^{m} \|\nabla f_j(\hat{x}) - \nabla f_j(\hat{x}_0)\| \\
&= \|\hat{z}\| \sum_{j=1}^{m} \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0) \right)^2}.
\end{aligned}$$

De lo anterior se desprende que, si

$$\left| \frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0) \right| < \varepsilon'$$

para toda $i \in \{1, \ldots, n\}$ y toda $j \in \{1, \ldots, m\}$, entonces

$$\begin{aligned}\|Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z})\| &\leq \|\hat{z}\| \sum_{j=1}^{m} \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0)\right)^2} \\ &< \|\hat{z}\| \sum_{j=1}^{m} \sqrt{\sum_{i=1}^{n} (\varepsilon')^2} \\ &= \|\hat{z}\| \sum_{j=1}^{m} \varepsilon' \sqrt{\sum_{i=1}^{n} 1} \\ &= \|\hat{z}\|\, \varepsilon' \sqrt{n} \sum_{j=1}^{m} 1 \\ &= \|\hat{z}\|\, m\sqrt{n}\varepsilon' \end{aligned} \tag{5.20}$$

para toda $\hat{z} \in \mathbb{R}^n$.

Esta última desigualdad lo que nos dice es que, si la distancia entre las entradas correspondientes de $Df(\hat{x})$ y $Df(\hat{x}_0)$ es "pequeña", entonces la distancia entre los valores de $Df(\hat{x})$ y $Df(\hat{x}_0)$ en cualquier $\hat{z} \in \mathbb{R}^n$ es "proporcionalmente pequeña" con respecto a la norma de $\hat{z}$.

Lo interesante de la condición anterior es que lo recíproco también es cierto. Es decir, si $\hat{x}, \hat{x}_0 \in U$ son tales que

$$\|Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z})\| \leq \|\hat{z}\|\,\varepsilon$$

para toda $\hat{z} \in \mathbb{R}^n$, entonces se tiene que

$$\left|\frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0)\right| \leq \varepsilon$$

para cada $i \in \{1, \ldots, n\}$ y cada $j \in \{1, \ldots, m\}$. En efecto, dado que

$$\begin{aligned} &Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z}) \\ &\qquad = ((\nabla f_1(\hat{x}) - \nabla f_1(\hat{x}_0)) \cdot \hat{z}, \ldots, (\nabla f_m(\hat{x}) - \nabla f_m(\hat{x}_0)) \cdot \hat{z}) \end{aligned}$$

para toda $\hat{z} \in \mathbb{R}^n$, entonces para cada $i \in \{1, \ldots, n\}$ se tiene que

$$Df(\hat{x})\,(\hat{e}_i) - Df(\hat{x}_0)\,(\hat{e}_i) = \left(\frac{\partial f_1}{\partial x_i}(\hat{x}) - \frac{\partial f_1}{\partial x_i}(\hat{x}_0), \ldots, \frac{\partial f_m}{\partial x_i}(\hat{x}) - \frac{\partial f_m}{\partial x_i}(\hat{x}_0)\right)$$

y por lo tanto

$$\begin{aligned}\left|\frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0)\right| &\leq \|Df(\hat{x})\,(\hat{e}_i) - Df(\hat{x}_0)\,(\hat{e}_i)\| \\ &\leq \|\hat{e}_i\|\,\varepsilon \\ &= \varepsilon\end{aligned}$$

para cada $j \in \{1, \ldots, m\}$.

La discusión anterior da lugar a la siguiente proposición, que nos aporta una caracterización de las funciones de clase C^1 en una región U.

Proposición 5.7 *Sea $f = (f_1, \ldots, f_m) : U \subset \mathbb{R}^n \to \mathbb{R}^m$. La función f es de clase C^1 en U si y sólo si*

1. *f es derivable para cada $\hat{x} \in U$, y*
2. *para cada $\hat{x}_0 \in U$ y cada $\varepsilon > 0$ existe $\delta > 0$ tal que $B_\delta(\hat{x}_0) \subset U$, y si $\hat{x} \in B_\delta(\hat{x}_0)$, entonces*

$$\|Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z})\| \leq \|\hat{z}\|\,\varepsilon$$

para toda $\hat{z} \in \mathbb{R}^n$.

Demostración.
$\Rightarrow$]
Como f es de clase C^1 en U, por la proposición 5.4 sabemos que f es derivable para cada $\hat{x} \in U$. Sean ahora $\hat{x}_0 \in U$ y $\varepsilon > 0$. Como f es de clase C^1 en U, entonces $\frac{\partial f_j}{\partial x_i}$ es continua en $\hat{x}_0$ para cada $i \in \{1, \ldots, n\}$ y cada $j \in \{1, \ldots, m\}$, de tal forma que existe $\delta > 0$ tal que $B_\delta(\hat{x}_0) \subset U$ y si $\hat{x} \in B_\delta(\hat{x}_0) \subset U$, entonces

$$\left|\frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0)\right| < \frac{1}{m\sqrt{n}}\varepsilon$$

para toda $i \in \{1, \ldots, n\}$ y toda $j \in \{1, \ldots, m\}$. Por tanto, tomando $\varepsilon' = \frac{1}{m\sqrt{n}}\varepsilon$ en la desigualdad 5.20, se tiene que

$$\begin{aligned} \|Df(\hat{x})(\hat{z}) - Df(\hat{x}_0)(\hat{z})\| &\leq \|\hat{z}\| \, m\sqrt{n}\varepsilon' \\ &= \|\hat{z}\| \, \varepsilon \end{aligned}$$

para toda $\hat{z} \in \mathbb{R}^n$.

$\Leftarrow]$

Recíprocamente, para $\hat{x}_0 \in U$ y $\varepsilon > 0$ sabemos que existe $\delta > 0$ tal que si $\hat{x} \in B_\delta(\hat{x}_0) \subset U$, entonces

$$\|Df(\hat{x})(\hat{z}) - Df(\hat{x}_0)(\hat{z})\| \leq \|\hat{z}\| \, \varepsilon$$

para toda $\hat{z} \in \mathbb{R}^n$.

Ahora, por la identidad 5.19, tomando $z = \hat{e}_i$ para cada $i \in \{1, \ldots, n\}$, se tiene que

$$\left(\frac{\partial f_1}{\partial x_i}(\hat{x}) - \frac{\partial f_1}{\partial x_i}(\hat{x}_0), \ldots, \frac{\partial f_m}{\partial x_i}(\hat{x}) - \frac{\partial f_m}{\partial x_i}(\hat{x}_0)\right) = Df(\hat{x})(\hat{e}_i) - Df(\hat{x}_0)(\hat{e}_i),$$

de modo que

$$\begin{aligned} \left|\frac{\partial f_j}{\partial x_i}(\hat{x}) - \frac{\partial f_j}{\partial x_i}(\hat{x}_0)\right| &\leq \left\| \left(\frac{\partial f_1}{\partial x_i}(\hat{x}) - \frac{\partial f_1}{\partial x_i}(\hat{x}_0), \ldots, \frac{\partial f_m}{\partial x_i}(\hat{x}) - \frac{\partial f_m}{\partial x_i}(\hat{x}_0)\right) \right\| \\ &= \|Df(\hat{x})(\hat{e}_i) - Df(\hat{x}_0)(\hat{e}_i)\| \\ &\leq \|\hat{e}_i\| \, \varepsilon \\ &= \varepsilon \end{aligned}$$

para cada $j \in \{1, \ldots, m\}$, lo que prueba que $\frac{\partial f_j}{\partial x_i}$ es continua en $\hat{x}_0 \in U$, y por lo tanto que f es de clase C^1 en U. ■

El tema principal de esta sección se centra en la búsqueda de condiciones para que una función f de $\mathbb{R}^n$ en $\mathbb{R}^n$ sea invertible en una vecindad de un punto $\hat{x}_0$. Dado que la derivabilidad de f en $\hat{x}_0$ nos garantiza que ésta se

parece mucho (en una vecindad de $\hat{x}_0$) a una función lineal de $\mathbb{R}^n$ en $\mathbb{R}^n$, la intuición nos dice que, si esta función lineal es invertible, entonces f también lo será (al menos en una vecindad del punto $\hat{x}_0$). Por lo anterior, empezaremos por dar condiciones bajo las cuales una función lineal L de $\mathbb{R}^n$ en $\mathbb{R}^n$ es invertible.

Antes de hacer esto, como seguramente el lector recordará de su curso de Álgebra Lineal, decimos que una función lineal $L : \mathbb{R}^n \to \mathbb{R}^n$ es invertible si existe otra función lineal $\tilde{L} : \mathbb{R}^n \to \mathbb{R}^n$ tal que $\left(L \circ \tilde{L}\right)(\hat{x}) = \hat{x} = \left(\tilde{L} \circ L\right)(\hat{x})$ para toda $\hat{x} \in \mathbb{R}^n$, y que cuando esta función lineal $\tilde{L}$ existe, entonces es única; por esta razón se le suele denotar por L^{-1}. También es oportuno tener presente que en este mismo curso de Álgebra Lineal se debieron haber probado dos condiciones necesarias y suficientes para que una función lineal $L : \mathbb{R}^n \to \mathbb{R}^n$ sea invertible. Una, que dice que L es invertible si y sólo si $L(\hat{x}) = \hat{0}$ sólo si $\hat{x} = \hat{0}$; y dos, que L es invertible si y sólo si la matriz M asociada a L (en cualesquiera bases de ambos $\mathbb{R}^n$, el dominio y el contradominio) es invertible, lo que a su vez es equivalente a que $\det(M) \neq 0$.

Una vez dicho lo anterior, lo siguiente que haremos será formular (apoyados en algunas de las condiciones mencionadas) otra condición necesaria y suficiente para que una función lineal $L : \mathbb{R}^n \to \mathbb{R}^n$ sea invertible.

Proposición 5.8 *Sea $L : \mathbb{R}^n \to \mathbb{R}^n$ una función lineal. L tiene inversa (o L es invertible) si y sólo si existe $m > 0$ tal que*

$$\|L(\hat{x})\| \geq m \|\hat{x}\|$$

para toda $\hat{x} \in \mathbb{R}^n$.

Demostración. Si L tiene inversa, sabemos que $L(\hat{x}) = \hat{0}$ sólo si $\hat{x} = \hat{0}$, de tal forma que, si consideramos el conjunto

$$S^{n-1} = \{\hat{x} \in \mathbb{R}^n \mid \|\hat{x}\| = 1\},$$

entonces $\|L(\hat{x})\| > 0$ para toda $\hat{x} \in S^{n-1}$.

Ahora, dado que S^{n-1} es un conjunto cerrado y acotado y L es una función continua (en $\mathbb{R}^n$), entonces $\|L\|$ alcanza un valor mínimo sobre S^{n-1}, es decir, existe $\hat{x}_0 \in S^{n-1}$ tal que

$$\|L(\hat{x})\| \geq \|L(\hat{x}_0)\| > 0$$

para toda $\hat{x} \in S^{n-1}$. Por tanto, si hacemos $m = \|L(\hat{x}_0)\| > 0$ y tomamos $\hat{x} \in \mathbb{R}^n$, con $\hat{x} \neq \hat{0}$ (pues claramente $\left\|L\left(\hat{0}\right)\right\| = 0 = m\left\|\hat{0}\right\|$), entonces

$$\frac{\hat{x}}{\|\hat{x}\|} \in S^{n-1}$$

y por lo tanto

$$L\left(\frac{\hat{x}}{\|\hat{x}\|}\right) \geq \|L(\hat{x}_0)\| = m,$$

de modo que

$$\|L(\hat{x})\| \geq m\|\hat{x}\|,$$

lo cual prueba la primera implicación.

Recíprocamente, si existe $m > 0$ tal que

$$\|L(\hat{x})\| \geq m\|\hat{x}\|$$

para toda $\hat{x} \in \mathbb{R}^n$, entonces se tiene que

$$\|L(\hat{x})\| \geq m\|\hat{x}\| > 0$$

para toda $\hat{x} \in \mathbb{R}^n$, con $\hat{x} \neq \hat{0}$, es decir que $L(\hat{x}) = \hat{0}$ sólo si $\hat{x} = \hat{0}$ lo que implica que L es invertible. ■

Con lo anterior ya tenemos las herramientas de Álgebra Lineal necesarias para probar el teorema principal de esta sección: el Teorema de la Función Inversa. Antes de hacer esto, e incluso antes de formular el teorema, daremos un par de lemas con los cuales iremos "preparando el terreno" para enunciarlo y probarlo. El primero de estos dos lemas nos permitirá probar que, si la derivada de una función f (de clase C^1 en un conjunto abierto U) es invertible en un punto $\hat{x}_0 \in U$ (en donde $Df(\hat{x}_0)$ se toma como función lineal o como matriz), entonces la derivada de f sigue siendo invertible para todo punto en una vecindad del punto $\hat{x}_0$.

Lema 5.1 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 en el conjunto abierto U, y $\hat{x}_0 \in U$. Si $Df(\hat{x}_0)$ es invertible, entonces existen $\delta > 0$ y $m > 0$ tales que $B_\delta(\hat{x}_0) \subset U$ y*

$$\|Df(\hat{x})\,(\hat{z})\| \geq m\,\|\hat{z}\|$$

para toda $\hat{x} \in B_\delta(\hat{x}_0)$ y para toda $\hat{z} \in \mathbb{R}^n$.

Demostración. Por la proposición 5.8, dado que $Df(\hat{x}_0)$ es invertible, sabemos que existe $m' > 0$ tal que

$$\|Df(\hat{x}_0)\,(\hat{z})\| \geq m'\,\|\hat{z}\|$$

para toda $\hat{z} \in \mathbb{R}^n$.

Por otra parte, como f es de clase C^1 en U (que es un abierto), por la proposición 5.7, tomando $\varepsilon = m'/2 > 0$, sabemos que existe $\delta > 0$ tal que $B_\delta(\hat{x}_0) \subset U$ y

$$\|Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z})\| \leq \|\hat{z}\|\,\frac{m'}{2}$$

para toda $\hat{x} \in B_\delta(\hat{x}_0)$ y para toda $\hat{z} \in \mathbb{R}^n$. Por lo tanto, de la desigualdad del triángulo tenemos que

$$\begin{aligned}\|Df(\hat{x})\,(\hat{z})\| &\geq \|Df(\hat{x}_0)\,(\hat{z})\| - \|Df(\hat{x})\,(\hat{z}) - Df(\hat{x}_0)\,(\hat{z})\| \\ &\geq m'\,\|\hat{z}\| - \|\hat{z}\|\,\frac{m'}{2} \\ &= \|\hat{z}\|\,\frac{m'}{2}\end{aligned}$$

para toda $\hat{x} \in B_\delta(\hat{x}_0) \subset U$ y para toda $\hat{z} \in \mathbb{R}^n$, de tal manera que tomando $m = m'/2 > 0$ logramos el resultado deseado. ■

Como una consecuencia inmediata de este lema y de la proposición 5.8, obtenemos el siguiente

Corolario 5.1 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 en U y $\hat{x}_0 \in U$. Si $Df(\hat{x}_0)$ es invertible, entonces existe $\delta > 0$ tal que $B_\delta(\hat{x}_0) \subset U$ y $Df(\hat{x})$ es invertible para toda $\hat{x} \in B_\delta(\hat{x}_0)$.*

El segundo lema que probaremos también es fundamental en la prueba del teorema de la función inversa, pues a partir de éste podremos asegurar que una función f es localmente invertible, y que la función inversa que se puede definir es continua en su dominio. Antes, sólo recordemos que en el capítulo 1 definimos otras normas para los elementos de $\mathbb{R}^n$, una de ellas llamada la norma infinito (que utilizaremos en la prueba del siguiente lema), y que está definida (y es denotada) como

$$\|\hat{x}\|_\infty := \text{máx}\{|x_1|, \ldots, |x_n|\}$$

para cada $\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, y que con relación a la norma euclideana satisface la desigualdad

$$\|\hat{x}\| \leq \sqrt{n}\, \|\hat{x}\|_\infty ,$$

la cual es válida para toda $\hat{x} \in \mathbb{R}^n$.

Lema 5.2 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 en el conjunto abierto U, y $\hat{x}_0 \in U$. Si $Df(\hat{x}_0)$ es invertible, entonces existen $\delta > 0$ y $m > 0$ tales que $B_\delta(\hat{x}_0) \subset U$ y*

$$m\, \|\hat{y} - \hat{x}\| \leq \|f(\hat{y}) - f(\hat{x})\| \tag{5.21}$$

para toda $\hat{x}, \hat{y} \in B_\delta(\hat{x}_0)$.

Demostración. Primero recordemos que si $f = (f_1, \ldots, f_n)$, entonces

$$Df(\hat{x})(\hat{z}) = (\nabla f_1(\hat{x}) \cdot \hat{z}, \ldots, \nabla f_n(\hat{x}) \cdot \hat{z})$$

para cada $\hat{x} \in U$ y cada $\hat{z} \in \mathbb{R}^n$.

Como $Df(\hat{x}_0)$ es invertible, por la proposición 5.8 sabemos que existe $m' > 0$ tal que

$$\|Df(\hat{x}_0)(\hat{z})\| \geq m' \|\hat{z}\|$$

para toda $\hat{z} \in \mathbb{R}^n$.

Por otra parte, por la proposición 5.7 sabemos que para $m'/(2\sqrt{n}) > 0$ existe $\delta > 0$ tal que si $\hat{x} \in B_\delta(\hat{x}_0) \subset U$, entonces

$$\|Df(\hat{x})(\hat{z}) - Df(\hat{x}_0)(\hat{z})\| \leq \|\hat{z}\| \frac{m'}{2\sqrt{n}}$$

para toda $\hat{z} \in \mathbb{R}^n$.

Probaremos que esta $\delta > 0$ y esta $m = m'/(2\sqrt{n}) > 0$ son las cantidades para las que se satisface la desigualdad 5.21, para toda $\hat{x}, \hat{y} \in B_\delta(\hat{x}_0) \subset U$.

Dados $\hat{x}, \hat{y} \in B_\delta(\hat{x}_0)$, sea $k \in \{1, \ldots, n\}$ tal que $\|Df(\hat{x}_0)(\hat{y} - \hat{x})\|_\infty = |\nabla f_k(\hat{x}_0) \cdot (\hat{y} - \hat{x})|$. Entonces

$$\begin{aligned} |\nabla f_k(\hat{x}_0) \cdot (\hat{y} - \hat{x})| &= \|Df(\hat{x}_0)(\hat{y} - \hat{x})\|_\infty \\ &\geq \frac{1}{\sqrt{n}} \|Df(\hat{x}_0)(\hat{y} - \hat{x})\|. \end{aligned}$$

Ahora, como $B_\delta(\hat{x}_0)$ es un conjunto convexo, sabemos que el segmento

$$[\hat{x}, \hat{y}] = \{\hat{x} + t(\hat{y} - \hat{x}) \in \mathbb{R}^n \mid t \in [0, 1]\}$$

está totalmente contenido en $B_\delta(\hat{x}_0) \subset U$, de tal forma que la función $\gamma_k : [0, 1] \subset \mathbb{R} \to \mathbb{R}$ definida como

$$\gamma_k(t) = f_k(\hat{x} + t(\hat{y} - \hat{x}))$$

es derivable para toda $t \in [0, 1]$.

De esta forma, aplicando el teorema del valor medio para funciones de $\mathbb{R}$ en $\mathbb{R}$, sabemos que existe $\xi_k \in (0, 1)$ tal que

$$\begin{aligned} f_k(\hat{y}) - f_k(\hat{x}) &= \gamma_k(1) - \gamma_k(0) \\ &= (1 - 0)\gamma_k'(\xi_k) \\ &= \gamma_k'(\xi_k) \\ &= \nabla f_k(\hat{x} + \xi_k(\hat{y} - \hat{x})) \cdot (\hat{y} - \hat{x}). \end{aligned}$$

Si recordamos ahora que para $\hat{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ se tiene que $|x_i| \leq \|\hat{x}\|$ para toda $i \in \{1, \ldots, n\}$, entonces también tenemos que

$$\|f(\hat{y}) - f(\hat{x})\| \geq |f_k(\hat{y}) - f_k(\hat{x})|$$

y

$$|\nabla f_k(\hat{x}) \cdot (\hat{y} - \hat{x}) - \nabla f_k(\hat{x}_0) \cdot (\hat{y} - \hat{x})| \leq \|Df(\hat{x}_0)(\hat{y} - \hat{x}) - Df(\hat{x})(\hat{y} - \hat{x})\|.$$

Con base en las desigualdades anteriores (y la desigualdad del triángulo), se tiene que

$$\begin{aligned}
&\|f(\hat{y}) - f(\hat{x})\| \\
&\geq |f_k(\hat{y}) - f_k(\hat{x})| \\
&= \left|\gamma_k'(\xi_k)\right| \\
&= |\nabla f_k(\hat{x} + \xi_k(\hat{y} - \hat{x})) \cdot (\hat{y} - \hat{x})| \\
&\geq |\nabla f_k(\hat{x}_0) \cdot (\hat{y} - \hat{x})| - |\nabla f_k(\hat{x} + \xi_k(\hat{y} - \hat{x})) \cdot (\hat{y} - \hat{x}) - \nabla f_k(\hat{x}_0) \cdot (\hat{y} - \hat{x})| \\
&\geq \frac{1}{\sqrt{n}} \|Df(\hat{x}_0)(\hat{y} - \hat{x})\| - \|Df(\hat{x} + \xi_k(\hat{y} - \hat{x}))(\hat{y} - \hat{x}) - Df(\hat{x}_0)(\hat{y} - \hat{x})\| \\
&\geq \frac{1}{\sqrt{n}} \left(m' \|\hat{y} - \hat{x}\|\right) - \frac{m'}{2\sqrt{n}} \|\hat{y} - \hat{x}\| \\
&= \frac{m'}{2\sqrt{n}} \|\hat{y} - \hat{x}\| \\
&= m \|\hat{y} - \hat{x}\|,
\end{aligned}$$

que es la desigualdad que se deseaba probar. ■

Como mencionamos anteriormente, de este lema se sigue el siguiente

Corolario 5.2 *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 en el conjunto abierto U, y $\hat{x}_0 \in U$. Si $Df(\hat{x}_0)$ es invertible, entonces existe $\delta > 0$ tal que $B_\delta(\hat{x}_0) \subset U$, f es inyectiva en $B_\delta(\hat{x}_0)$, y además*

$$f^{-1} : f(B_\delta(\hat{x}_0)) \subset \mathbb{R}^n \to \mathbb{R}^n$$

es uniformemente continua en su dominio.

Demostración. Por el lema anterior sabemos que existen $\delta > 0$ y $m > 0$ tales que

$$\left\|f(\hat{x}') - f(\hat{x})\right\| \geq m \left\|\hat{x}' - \hat{x}\right\|$$

para todo $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0) \subset U$.

Ahora, si tomamos $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0)$, con $\hat{x} \neq \hat{x}'$, entonces

$$\begin{aligned} \left\| f\left(\hat{x}'\right) - f\left(\hat{x}\right) \right\| &\geq m \left\| \hat{x}' - \hat{x} \right\| \\ &> 0, \end{aligned}$$

de modo que $f\left(\hat{x}'\right) \neq f\left(\hat{x}\right)$ y por lo tanto se tiene que f es inyectiva en $B_\delta(\hat{x}_0)$.

Por otra parte, dado $\varepsilon > 0$, si tomamos $\delta' = \varepsilon m$ y $\hat{y}, \hat{y}' \in f\left(B_\delta(\hat{x}_0)\right)$ tales que $\left\| \hat{y}' - \hat{y} \right\| < \delta'$, si $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0)$ son tales que $\hat{y}' = f\left(\hat{x}'\right)$ y $\hat{y} = f\left(\hat{x}\right)$, se tiene que

$$\begin{aligned} \left\| f^{-1}\left(\hat{y}'\right) - f^{-1}\left(\hat{y}\right) \right\| &= \left\| \hat{x}' - \hat{x} \right\| \\ &\leq \frac{1}{m} \left\| f\left(\hat{x}'\right) - f\left(\hat{x}\right) \right\| \\ &= \frac{1}{m} \left\| \hat{y}' - \hat{y} \right\| \\ &< \frac{1}{m} \delta' \\ &= \varepsilon, \end{aligned}$$

lo que prueba que f^{-1} es uniformemente continua (y por tanto continua) en $f\left(B_\delta(\hat{x}_0)\right)$. ■

Con base en todo el trabajo realizado previamente ya estamos en condiciones de formular y probar el teorema de la función inversa.

Teorema 5.3 (de la función inversa) *Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 en el conjunto abierto U, y $\hat{x}_0 \in U$. Si $Df(\hat{x}_0)$ es invertible, entonces existe $\delta > 0$ tal que:*

1. *$B_\delta(\hat{x}_0) \subset U$ y f es inyectiva en $B_\delta(\hat{x}_0)$,*
2. *$f^{-1} : f\left(B_\delta(\hat{x}_0)\right) \subset \mathbb{R}^n \to \mathbb{R}^n$ es continua en $f\left(B_\delta(\hat{x}_0)\right)$,*
3. *$f\left(B_\delta(\hat{x}_0)\right) \subset \mathbb{R}^n$ es un conjunto abierto, y*

4. *f^{-1} es de clase C^1 en $f(B_\delta(\hat{x}_0))$ y además, si $\hat{y} = f(\hat{x}) \in f(B_\delta(\hat{x}_0))$, entonces*
$$Df^{-1}(\hat{y}) = Df^{-1}(f(\hat{x})) = (Df(\hat{x}))^{-1}.$$

Demostración. Por los lemas 5.1 y 5.2 sabemos que existen $\delta > 0$ y $m > 0$ tales que
$$m\left\|\hat{x}' - \hat{x}\right\| \leq \left\|f(\hat{x}') - f(\hat{x})\right\| \tag{5.22}$$
para toda $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0) \subset U$, y que
$$\|Df(\hat{x})(\hat{z})\| \geq m\|\hat{z}\| \tag{5.23}$$
para toda $\hat{x} \in B_\delta(\hat{x}_0)$ y toda $\hat{z} \in \mathbb{R}^n$.

Ahora, por el corolario 5.2 se tiene que f es inyectiva en $B_\delta(\hat{x}_0)$ y que f^{-1} es continua en $f(B_\delta(\hat{x}_0))$, con lo cual se tiene la prueba de los incisos 1 y 2 del enunciado.

Probaremos ahora que $f(B_\delta(\hat{x}_0)) \subset \mathbb{R}^n$ es un conjunto abierto. Sea $\hat{y}' = f(\hat{x}') \in f(B_\delta(\hat{x}_0))$, con $\hat{x}' \in B_\delta(\hat{x}_0)$. Como $B_\delta(\hat{x}_0)$ es un conjunto abierto, existe $\delta' > 0$ tal que el conjunto
$$A = \left\{\hat{x} \in \mathbb{R}^n \mid \left\|\hat{x} - \hat{x}'\right\| \leq \delta'\right\}$$
se queda contenido en $B_\delta(\hat{x}_0)$. Probaremos que existe $r > 0$ tal que $B_r(\hat{y}') \subset f(A) \subset f(B_\delta(\hat{x}_0))$.

Para "justificar" de manera intuitiva el valor de r que vamos a tomar, obsérvese que si $\hat{y}'' \in B_r(\hat{y}')$ fuera tal que $\hat{y}'' = f(\hat{x})$ para alguna $\hat{x} \in A$, por la desigualdad 5.22 se debería tener que
$$\begin{aligned}\left\|\hat{x} - \hat{x}'\right\| &= \left\|f^{-1}(\hat{y}'') - f^{-1}(\hat{y}')\right\| \\ &\leq \frac{1}{m}\left\|\hat{y}'' - \hat{y}'\right\| \\ &< \frac{r}{m},\end{aligned}$$
de tal forma que, para no caer en contradicción con el supuesto de que $\hat{x} \in A$, se deberá elegir r de tal forma que $r/m \leq \delta'$. Con base en el razonamiento

anterior, tomamos $r = m\delta'/2 > 0$ y $\hat{y}'' \in B_r(\hat{y}')$. Para probar que existe $\hat{x}'' \in A$ tal que $f(\hat{x}'') = \hat{y}''$, definimos $h : U \subset \mathbb{R}^n \to \mathbb{R}$ como

$$\begin{aligned} h(\hat{x}) &= \left\| f(\hat{x}) - \hat{y}'' \right\|^2 \\ &= \left(f(\hat{x}) - \hat{y}'' \right) \cdot \left(f(\hat{x}) - \hat{y}'' \right). \end{aligned}$$

Dado que $A \subset U$ es un conjunto cerrado y acotado, y h es continua en U, sabemos que h alcanza un valor mínimo sobre A, es decir, que existe $\hat{x}'' \in A$ tal que

$$\begin{aligned} \left\| f(\hat{x}'') - \hat{y}'' \right\|^2 &= h(\hat{x}'') \\ &\leq h(\hat{x}) \\ &= \left\| f(\hat{x}) - \hat{y}'' \right\|^2 \end{aligned}$$

para toda $\hat{x} \in A$. Nótese que ahora nuestro objetivo es probar que $h(\hat{x}'') = 0$, pues de este hecho se concluye que $f(\hat{x}'') = \hat{y}''$.

Ahora, como $\hat{x}'' \in A$, se tiene que $\|\hat{x}'' - \hat{x}'\| \leq \delta'$, de modo que, aún cuando h alcanza un valor mínimo en $\hat{x}''$, no podemos asegurar que éste sea un punto crítico de h.

Con el fin de probar que $\hat{x}''$ sí es un punto crítico de h, descartaremos la posibilidad de que $\|\hat{x}'' - \hat{x}'\| = \delta'$. Si este fuera el caso, por la desigualdad del triángulo se tendría que

$$\begin{aligned} \sqrt{h(\hat{x}'')} &= \left\| f(\hat{x}'') - \hat{y}'' \right\| \\ &\geq \left\| f(\hat{x}'') - \hat{y}' \right\| - \left\| \hat{y}' - \hat{y}'' \right\| \\ &= \left\| f(\hat{x}'') - f(\hat{x}') \right\| - \left\| \hat{y}' - \hat{y}'' \right\| \\ &\geq m \left\| \hat{x}'' - \hat{x}' \right\| - \left\| \hat{y}' - \hat{y}'' \right\| \\ &> m\delta' - \frac{m\delta'}{2} \\ &= \frac{m\delta'}{2} \\ &> \left\| \hat{y}' - \hat{y}'' \right\| \\ &= \left\| f(\hat{x}') - \hat{y}'' \right\| \\ &= \sqrt{h(\hat{x}')}, \end{aligned}$$

lo que contradice el hecho de que $h\left(\hat{x}''\right)$ es el valor mínimo de h sobre A. De esta forma, se debe tener que $\|\hat{x}'' - \hat{x}'\| < \delta'$ y por tanto $\hat{x}''$ es un punto crítico de h, es decir que

$$\begin{aligned} Dh\left(\hat{x}''\right) &= 2\left(f\left(\hat{x}''\right) - \hat{y}''\right) \cdot Df\left(\hat{x}''\right) \\ &= \hat{0} \end{aligned}$$

Si ahora recordamos que, por la desigualdad 5.23 y la proposición 5.8, se tiene que $Df\left(\hat{x}\right)$ es invertible (y por tanto suprayectiva) para toda $\hat{x} \in B_\delta(\hat{x}_0)$, entonces debe existir $\hat{z} \in \mathbb{R}^n$ tal que

$$Df\left(\hat{x}''\right)\left(\hat{z}\right) = f\left(\hat{x}''\right) - \hat{y}'',$$

de tal forma que en particular se tiene que

$$\begin{aligned} 0 &= \left(\left(f\left(\hat{x}''\right) - \hat{y}''\right) \cdot Df\left(\hat{x}''\right)\right)\left(\hat{z}\right) \\ &= \left(f\left(\hat{x}''\right) - \hat{y}''\right) \cdot \left(Df\left(\hat{x}''\right)\left(\hat{z}\right)\right) \\ &= \left(f\left(\hat{x}''\right) - \hat{y}''\right) \cdot \left(f\left(\hat{x}''\right) - \hat{y}''\right) \\ &= \left\|f\left(\hat{x}''\right) - \hat{y}''\right\|^2 \end{aligned}$$

y por lo tanto que $f\left(\hat{x}''\right) = \hat{y}''$, lo que prueba que $B_r(\hat{y}') \subset f\left(B_\delta(\hat{x}_0)\right)$, es decir que $f\left(B_\delta(\hat{x}_0)\right)$ es un conjunto abierto.

Para probar el último inciso, nuestro primer paso será demostrar que f^{-1} es derivable en todo punto $\hat{y}' = f(\hat{x}') \in f\left(B_\delta(\hat{x}_0)\right)$. Usando nuevamente que $Df(\hat{x}')$ es invertible para toda $\hat{x}' \in B_\delta(\hat{x}_0)$, sólo nos restará mostrar que la función lineal

$$L = \left(Df(\hat{x}')\right)^{-1}$$

satisface la definición 5.1, es decir, que

$$\lim_{\hat{y} \to \hat{y}'} \frac{f^{-1}(\hat{y}) - f^{-1}(\hat{y}') - L(\hat{y} - \hat{y}')}{\|\hat{y} - \hat{y}'\|} = \hat{0}. \tag{5.24}$$

Para ello, recurriremos a la función $g : B_\delta(\hat{x}_0) \subset \mathbb{R}^n \to \mathbb{R}^n$ dada por

$$g\left(\hat{x}\right) = \begin{cases} \frac{\hat{x} - \hat{x}' - L(f(\hat{x}) - f(\hat{x}'))}{\|f(\hat{x}) - f(\hat{x}')\|} & \text{si } \hat{x} \neq \hat{x}' \\ \\ \hat{0} & \text{si } \hat{x} = \hat{x}' \end{cases}$$

la cual está bien definida, puesto que f es inyectiva en la bola $B_\delta(\hat{x}_0)$.

Obsérvese que la composición $g \circ f^{-1}$ también está bien definida sobre el conjunto abierto $f\left(B_\delta(\hat{x}_0)\right) \setminus \{\hat{y}'\}$ y que

$$\begin{aligned}\left(g \circ f^{-1}\right)(\hat{y}) &= g\left(f^{-1}(\hat{y})\right) \\ &= \begin{cases} \frac{f^{-1}(\hat{y}) - f^{-1}(\hat{y}') - L(\hat{y}-\hat{y}')}{\|\hat{y}-\hat{y}'\|} & \text{si } \hat{y} \neq \hat{y}' \\ \\ \hat{0} & \text{si } \hat{y} = \hat{y}' \end{cases}\end{aligned}$$

para toda $\hat{y} \in f\left(B_\delta(\hat{x}_0)\right) \setminus \{\hat{y}'\}$.

En virtud de lo anterior, tenemos que

$$\lim_{\hat{y}\to\hat{y}'} \frac{f^{-1}(\hat{y}) - f^{-1}(\hat{y}') - L(\hat{y}-\hat{y}')}{\|\hat{y}-\hat{y}'\|} = \lim_{\hat{y}\to\hat{y}'} \left(g \circ f^{-1}\right)(\hat{y}),$$

de tal forma que para obtener la identidad 5.24, como f^{-1} es continua en $\hat{y}'$, por el inciso 6 de la proposición 2.12 del capítulo 2, bastará mostrar que g es continua en $\hat{x}' = f^{-1}(\hat{y}')$, lo que es equivalente a probar que

$$\begin{aligned}\lim_{\hat{x}\to\hat{x}'} g(\hat{x}) &= \lim_{\hat{x}\to\hat{x}'} \frac{\hat{x} - \hat{x}' - \left(Df(\hat{x}')\right)^{-1}\left(f(\hat{x}) - f(\hat{x}')\right)}{\|f(\hat{x}) - f(\hat{x}')\|} \\ &= \hat{0}.\end{aligned}$$

Para probar este último límite, nótese que

$$\begin{aligned}&\frac{\hat{x} - \hat{x}' - L(f(\hat{x}) - f(\hat{x}'))}{\|f(\hat{x}) - f(\hat{x}')\|} \\ &= \frac{\hat{x} - \hat{x}' - \left(Df(\hat{x}')\right)^{-1}\left(f(\hat{x}) - f(\hat{x}')\right)}{\|f(\hat{x}) - f(\hat{x}')\|} \\ &= \frac{\left(Df(\hat{x}')\right)^{-1}\left[Df(\hat{x}')\left(\hat{x} - \hat{x}'\right) - \left(f(\hat{x}) - f(\hat{x}')\right)\right]}{\|f(\hat{x}) - f(\hat{x}')\|} \\ &= -\frac{\|\hat{x} - \hat{x}'\|}{\|f(\hat{x}) - f(\hat{x}')\|}\left(Df(\hat{x}')\right)^{-1}\left(\frac{f(\hat{x}) - f(\hat{x}') - Df(\hat{x}')\left(\hat{x} - \hat{x}'\right)}{\|\hat{x} - \hat{x}'\|}\right)\end{aligned}$$

de tal forma que, como $(Df(\hat{x}'))^{-1}$ es una función continua (toda función lineal es continua) y f es derivable en $\hat{x}'$, se tiene que

$$\begin{aligned}
&\lim_{\hat{x}\to\hat{x}'} \left(Df(\hat{x}')\right)^{-1} \left(\frac{f(\hat{x}) - f(\hat{x}') - Df(\hat{x}')\,(\hat{x}-\hat{x}')}{\|\hat{x}-\hat{x}'\|}\right) \\
&= \left(Df(\hat{x}')\right)^{-1} \left(\lim_{\hat{x}\to\hat{x}'} \frac{f(\hat{x}) - f(\hat{x}') - Df(\hat{x}')\,(\hat{x}-\hat{x}')}{\|\hat{x}-\hat{x}'\|}\right) \\
&= \left(Df(\hat{x}')\right)^{-1} \left(\hat{0}\right) \\
&= \hat{0}.
\end{aligned}$$

Ahora, como por la desigualdad 5.4 sabemos que

$$\frac{\|\hat{x}-\hat{x}'\|}{\|f(\hat{x}) - f(\hat{x}')\|} \leq \frac{1}{m}$$

para toda $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0)$, con $\hat{x} \neq \hat{x}'$, entonces

$$\begin{aligned}
&\lim_{\hat{x}\to\hat{x}'} \frac{\hat{x}-\hat{x}' - (Df(\hat{x}'))^{-1}\,(f(\hat{x}) - f(\hat{x}'))}{\|f(\hat{x}) - f(\hat{x}')\|} \\
&= -\lim_{\hat{x}\to\hat{x}'} \frac{\|\hat{x}-\hat{x}'\|}{\|f(\hat{x}) - f(\hat{x}')\|} \left(Df(\hat{x}')\right)^{-1} \left(\frac{f(\hat{x}) - f(\hat{x}') - Df(\hat{x}')\,(\hat{x}-\hat{x}')}{\|\hat{x}-\hat{x}'\|}\right) \\
&= \hat{0}.
\end{aligned}$$

Lo anterior prueba la identidad 5.24. Por lo tanto f^{-1} es derivable en $\hat{y}' = f(\hat{x}') \in f\left(B_\delta(\hat{x}_0)\right)$, y además

$$\begin{aligned}
Df^{-1}\left(\hat{y}'\right) &= Df^{-1}\left(f(\hat{x}')\right) \\
&= \left(Df(\hat{x}')\right)^{-1} \\
&= \left(Df(f^{-1}(\hat{y}'))\right)^{-1}
\end{aligned}$$

para toda $\hat{y}' \in f\left(B_\delta(\hat{x}_0)\right)$.

Finalmente, probaremos que f^{-1} es de clase C^1 en $f\left(B_\delta(\hat{x}_0)\right)$ y para ello usaremos el criterio de la proposición 5.7. Antes, observemos que de la desigualdad 5.23 y el hecho de que

$$Df^{-1}\left(f(\hat{x})\right) = \left(Df(\hat{x})\right)^{-1} \tag{5.25}$$

para toda $\hat{x} \in B_\delta(\hat{x}_0)$, se tiene que

$$\begin{aligned}\|\hat{z}\| &= \left\| Df(\hat{x}) \left((Df(\hat{x}))^{-1} (\hat{z}) \right) \right\| \\ &\geq m \left\| (Df(\hat{x}))^{-1} (\hat{z}) \right\|\end{aligned}$$

y por lo tanto

$$\begin{aligned}\left\| Df^{-1} (f(\hat{x})) (\hat{z}) \right\| &= \left\| (Df(\hat{x}))^{-1} (\hat{z}) \right\| \\ &\leq \frac{1}{m} \|\hat{z}\| \end{aligned} \tag{5.26}$$

para toda $\hat{z} \in \mathbb{R}^n$.

También notemos que, si $A, B \in \mathbb{R}$ son diferentes de 0, entonces

$$\begin{aligned}A^{-1} - B^{-1} &= \frac{1}{A} - \frac{1}{B} \\ &= \frac{B - A}{AB} \\ &= \frac{1}{A} (B - A) \frac{1}{B} \\ &= A^{-1} (B - A) B^{-1}\end{aligned}$$

es decir, que

$$A^{-1} - B^{-1} = A^{-1} (B - A) B^{-1}.$$

Ahora observe que esta última identidad sigue siendo válida si A y B son matrices de $n \times n$ invertibles, de modo que si $L_1, L_2 : \mathbb{R}^n \to \mathbb{R}^n$ son funciones lineales que tienen inversa, entonces también se cumple que

$$L_1^{-1} - L_2^{-1} = L_1^{-1} \circ [L_2 - L_1] \circ L_2^{-1}$$

De esta última identidad y usando 5.25 para $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0)$, se deduce que

$$\begin{aligned}&Df^{-1} (f(\hat{x})) - Df^{-1} \left(f(\hat{x}') \right) \\ &\quad = (Df(\hat{x}))^{-1} - \left(Df(\hat{x}') \right)^{-1} \\ &\quad = (Df(\hat{x}))^{-1} \circ \left[Df(\hat{x}') - Df(\hat{x}) \right] \circ \left(Df(\hat{x}') \right)^{-1} \\ &\quad = Df^{-1} (f(\hat{x})) \circ \left[\left(Df \left(\hat{x}' \right) - Df (\hat{x}) \right) \right] \circ Df^{-1} \left(f(\hat{x}') \right)\end{aligned}$$

y por la desigualdad 5.26, se tiene que

$$\begin{aligned}
&\left\|Df^{-1}(f(\hat{x}))(\hat{z}) - Df^{-1}(f(\hat{x}'))(\hat{z})\right\| \\
&= \left\|(Df^{-1}(f(\hat{x})) \circ [Df(\hat{x}') - Df(\hat{x})] \circ Df^{-1}(f(\hat{x}')))(\hat{z})\right\| \\
&= \left\|Df^{-1}\left(f(\hat{x})\right)\left[\left[Df\left(\hat{x}'\right) - Df\left(\hat{x}\right)\right]\left(Df^{-1}\left(f(\hat{x}')\right)\left(\hat{z}\right)\right)\right]\right\| \\
&\leq \frac{1}{m}\left\|\left[Df\left(\hat{x}'\right) - Df\left(\hat{x}\right)\right]\left(Df^{-1}\left(f(\hat{x}')\right)\left(\hat{z}\right)\right)\right\| \qquad (5.27)
\end{aligned}$$

para todas $\hat{x}, \hat{x}' \in B_\delta(\hat{x}_0)$ y toda $\hat{z} \in \mathbb{R}^n$.

Una vez dicho lo anterior, sean $\hat{y}' = f(\hat{x}') \in f\left(B_\delta(\hat{x}_0)\right)$ y $\varepsilon > 0$. Como f es de clase C^1 en $B_\delta(\hat{x}_0)$, por la proposición 5.7 ($\Longrightarrow$), y el hecho de que $B_\delta(\hat{x}_0)$ es un abierto, sabemos que para la cantidad $m^2\varepsilon > 0$ existe $\varepsilon' > 0$ tal que $B_{\varepsilon'}(\hat{x}') \subset B_\delta(\hat{x}_0)$, y si $\hat{x} \in B_{\varepsilon'}(\hat{x}')$, entonces

$$\left\|\left[Df\left(\hat{x}'\right) - Df\left(\hat{x}\right)\right]\left(\hat{z}\right)\right\| \leq m^2\varepsilon\left\|\hat{z}\right\| \qquad (5.28)$$

para toda $\hat{z} \in \mathbb{R}^n$.

Finalmente, como f^{-1} es continua en $f\left(B_\delta(\hat{x}_0)\right)$ (que ya probamos que es un conjunto abierto), sabemos que para $\varepsilon' > 0$ existe $\delta' > 0$ tal que $B_{\delta'}(\hat{y}') \subset f\left(B_\delta(\hat{x}_0)\right)$ y si $\hat{y} = f(\hat{x}) \in B_{\delta'}(\hat{y}')$, entonces

$$\left\|f^{-1}\left(\hat{y}\right) - f^{-1}\left(\hat{y}'\right)\right\| = \left\|\hat{x} - \hat{x}'\right\| < \varepsilon'. \qquad (5.29)$$

Por lo tanto, aplicando las desigualdades 5.27, 5.28 y 5.26, concluimos que

$$\begin{aligned}
&\left\|Df^{-1}\left(\hat{y}\right)\left(\hat{z}\right) - Df^{-1}\left(\hat{y}'\right)\left(\hat{z}\right)\right\| \\
&\qquad = \left\|Df^{-1}\left(f(\hat{x})\right)\left(\hat{z}\right) - Df^{-1}\left(f(\hat{x}')\right)\left(\hat{z}\right)\right\| \\
&\qquad \leq \frac{1}{m}\left\|\left[Df\left(\hat{x}'\right) - Df\left(\hat{x}\right)\right]\left(Df^{-1}\left(f(\hat{x}')\right)\left(\hat{z}\right)\right)\right\| \\
&\qquad \leq \frac{1}{m}\left(m^2\varepsilon\left\|Df^{-1}\left(f(\hat{x}')\right)\left(\hat{z}\right)\right\|\right) \\
&\qquad = m\varepsilon\left\|Df^{-1}\left(f(\hat{x}')\right)\left(\hat{z}\right)\right\| \\
&\qquad \leq m\varepsilon\left(\frac{1}{m}\left\|\hat{z}\right\|\right) \\
&\qquad = \left\|\hat{z}\right\|\varepsilon
\end{aligned}$$

para toda $\hat{y} = f(\hat{x}) \in B_{\delta'}(\hat{y}')$ y para toda $\hat{z} \in \mathbb{R}^n$, de tal forma que nuevamente por la proposición 5.7 ($\Longleftarrow$) tenemos que f^{-1} es de clase C^1 en $f\left(B_\delta(\hat{x}_0)\right)$, con lo cual terminamos la prueba. ■

El teorema de la función inversa es un resultado de caracter teórico muy importante, pero también lo es desde un punto de vista práctico. Como prueba de ello, mostraremos cómo se emplea en el siguiente problema de cambio de coordenadas: si una función f de $\mathbb{R}^3$ en $\mathbb{R}$ está dada en términos de coordenadas esféricas, nuestro problema será encontrar la expresión de su derivada en términos de las coordenadas cartesianas, sin necesidad de escribir a f en términos de éstas últimas.

Ejemplo 5.4 *Sea* $g : \mathbb{R}^3 \to \mathbb{R}^3$ *definida como*

$$g\left(\rho, \theta, \varphi\right) = \left(\rho \operatorname{sen}(\varphi) \cos(\theta), \rho \operatorname{sen}(\varphi) \operatorname{sen}(\theta), \rho \cos(\varphi)\right).$$

Seguramente el lector estará de acuerdo en que esta función es la función de cambio de coordenadas esféricas a coordenadas cartesianas *para puntos de* $\mathbb{R}^3$.

También podemos concluir con toda certeza que g *es de clase* C^k *en* $\mathbb{R}^3$ *para toda* $k \in \mathbb{N}$, *y por lo tanto derivable en todo punto de* $\mathbb{R}^3$, *con*

$$Dg\left(\rho, \theta, \varphi\right) = \begin{bmatrix} \operatorname{sen}(\varphi)\cos(\theta) & -\rho \operatorname{sen}(\varphi)\operatorname{sen}(\theta) & \rho\cos(\varphi)\cos(\theta) \\ \operatorname{sen}(\varphi)\operatorname{sen}(\theta) & \rho\operatorname{sen}(\varphi)\cos(\theta) & \rho\cos(\varphi)\operatorname{sen}(\theta) \\ \cos(\varphi) & 0 & -\rho\operatorname{sen}(\varphi) \end{bmatrix}.$$

Si ahora observamos que

$$\det\left(Dg\left(\rho, \theta, \varphi\right)\right) = -\rho^2 \operatorname{sen}(\varphi),$$

tendremos que $Dg\left(\rho, \theta, \varphi\right)$ *es invertible para toda* $(\rho, \theta, \varphi) \in \mathbb{R}^3$, *si* $\rho \neq 0$ *y* $\varphi \neq k\pi$, *con* $k \in \mathbb{Z}$, *y su inversa (calculada por el método que mejor conozca el lector) estará dada por*

$$\left(Dg\left(\rho, \theta, \varphi\right)\right)^{-1} = \begin{bmatrix} \operatorname{sen}(\varphi)\cos(\theta) & \operatorname{sen}(\varphi)\operatorname{sen}(\theta) & \cos(\varphi) \\ -\frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)} & \frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)} & 0 \\ \frac{1}{\rho}\cos(\theta)\cos(\varphi) & \frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta) & -\frac{1}{\rho}\operatorname{sen}(\varphi) \end{bmatrix}.$$

De esta forma, el teorema de la función inversa nos asegura que, si $(\rho_0, \theta_0, \varphi_0) \in \mathbb{R}^3$ *es tal que* $\rho_0 \neq 0$ *y* $k\pi < \varphi_0 < (k+1)\pi$ *para alguna* $k \in \mathbb{Z}$*, entonces existe* $\delta > 0$ *tal que la función* g *es inyectiva en la bola* $B_\delta(\rho_0, \theta_0, \varphi_0)$*, que la función* g^{-1} *está definida en el abierto* $g(B_\delta(\rho_0, \theta_0, \varphi_0))$ *y que para toda* $(\rho, \theta, \varphi) \in B_\delta(\rho_0, \theta_0, \varphi_0)$ *se tiene que*

$$Dg^{-1}(g(\rho,\theta,\varphi)) = (Dg(\rho,\theta,\varphi))^{-1}$$
$$= \begin{bmatrix} \operatorname{sen}(\varphi)\cos(\theta) & \operatorname{sen}(\varphi)\operatorname{sen}(\theta) & \cos(\varphi) \\ -\frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)} & \frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)} & 0 \\ \frac{1}{\rho}\cos(\theta)\cos(\varphi) & \frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta) & -\frac{1}{\rho}\operatorname{sen}(\varphi) \end{bmatrix}.$$

Dado que g *es la función de cambio de coordenadas esféricas a coordenadas cartesianas, entonces* g^{-1} *será la función de cambio de coordenadas cartesianas a coordenadas esféricas, función que no es tan sencilla de calcular explícitamente. Pero aún cuando no tengamos una expresión explícita para* g^{-1}*, el hecho de que tengamos una expresión para su derivada nos resulta de mucha ayuda. En efecto, si* $f : U \subset \mathbb{R}^3 \to \mathbb{R}$ *es una función de clase* C^1 *la cual está dada en términos de coordenadas esféricas, de la misma forma que hicimos en la subsección 5.3.1 con la función de cambio de coordenadas polares a cartesianas, podemos asumir que la función* $f \circ g^{-1}$ *es la misma función* f *sólo que expresada en términos de coordenadas cartesianas.*

Por tanto, recurriendo nuevamente a la regla de la cadena, tendremos que

$$\begin{aligned} \nabla\left(f \circ g^{-1}\right)(x,y,z) &= D\left(f \circ g^{-1}\right)(x,y,z) \\ &= Df\left(g^{-1}(x,y,z)\right) Dg^{-1}(x,y,z) \\ &= \nabla f\left(g^{-1}(x,y,z)\right) Dg^{-1}(x,y,z) \end{aligned}$$

de donde concluimos que

$$\begin{aligned} &\frac{\partial\left(f \circ g^{-1}\right)}{\partial x}(x,y,z) \\ &= \operatorname{sen}(\varphi)\cos(\theta)\frac{\partial f}{\partial \rho}\left(g^{-1}(x,y,z)\right) - \frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial \theta}\left(g^{-1}(x,y,z)\right) \end{aligned}$$

$$
\begin{aligned}
&\quad + \frac{1}{\rho}\cos(\theta)\cos(\varphi)\frac{\partial f}{\partial \varphi}\left(g^{-1}(x,y,z)\right)\\
&= \operatorname{sen}(\varphi)\cos(\theta)\left(\frac{\partial f}{\partial \rho}\circ g^{-1}\right)(x,y,z) - \frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)}\left(\frac{\partial f}{\partial \theta}\circ g^{-1}\right)(x,y,z)\\
&+ \frac{1}{\rho}\cos(\theta)\cos(\varphi)\left(\frac{\partial f}{\partial \varphi}\circ g^{-1}\right)(x,y,z)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial\left(f\circ g^{-1}\right)}{\partial y}(x,y,z)\\
&= \operatorname{sen}(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial \rho}\left(g^{-1}(x,y,z)\right) + \frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial \theta}\left(g^{-1}(x,y,z)\right)\\
&\quad + \frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial \varphi}\left(g^{-1}(x,y,z)\right)\\
&= \operatorname{sen}(\varphi)\operatorname{sen}(\theta)\left(\frac{\partial f}{\partial \rho}\circ g^{-1}\right)(x,y,z) + \frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)}\left(\frac{\partial f}{\partial \theta}\circ g^{-1}\right)(x,y,z)\\
&+ \frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta)\left(\frac{\partial f}{\partial \varphi}\circ g^{-1}\right)(x,y,z)
\end{aligned}
$$

y

$$
\begin{aligned}
&\frac{\partial\left(f\circ g^{-1}\right)}{\partial z}(x,y,z)\\
&\qquad = \cos(\varphi)\left(\frac{\partial f}{\partial \rho}\circ g^{-1}\right)(x,y,z) - \frac{1}{\rho}\operatorname{sen}(\varphi)\left(\frac{\partial f}{\partial \varphi}\circ g^{-1}\right)(x,y,z).
\end{aligned}
$$

Nótese que las identidades anteriores también se pueden escribir en términos de las coordenadas esféricas (ρ,θ,φ) *(que son de las que estamos suponiendo que depende* f*). En efecto, si recordamos que* $(x,y,z) = g(\rho,\theta,\varphi)$*, (o que* $(\rho,\theta,\varphi) = g^{-1}(x,y,z)$*) se tiene que*

$$\frac{\partial\left(f\circ g^{-1}\right)}{\partial x}(x,y,z)=\operatorname{sen}(\varphi)\cos(\theta)\frac{\partial f}{\partial\rho}(\rho,\theta,\varphi)-\frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial\theta}(\rho,\theta,\varphi)+\frac{1}{\rho}\cos(\theta)\cos(\varphi)\frac{\partial f}{\partial\varphi}(\rho,\theta,\varphi)$$

$$\frac{\partial\left(f\circ g^{-1}\right)}{\partial y}(x,y,z)=\operatorname{sen}(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial\rho}(\rho,\theta,\varphi)+\frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial\theta}(\rho,\theta,\varphi)+\frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial\varphi}(\rho,\theta,\varphi)$$

y

$$\frac{\partial\left(f\circ g^{-1}\right)}{\partial z}(x,y,z)=\cos(\varphi)\frac{\partial f}{\partial\rho}(\rho,\theta,\varphi)-\frac{1}{\rho}\operatorname{sen}(\varphi)\frac{\partial f}{\partial\varphi}(\rho,\theta,\varphi).$$

Si no usamos ningún tipo de coordenadas para el punto $\hat{x}$, y cometemos un abuso de notación escribiendo que $g^{-1}(\hat{x})=\hat{x}$, entonces $f=f\circ g^{-1}$, de modo que las identidades anteriores se pueden escribir como

$$\begin{aligned}\frac{\partial f}{\partial x}(\hat{x})&=\frac{\partial\left(f\circ g^{-1}\right)}{\partial x}(\hat{x})\\&=\operatorname{sen}(\varphi)\cos(\theta)\frac{\partial f}{\partial\rho}(\hat{x})-\frac{\operatorname{sen}(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial\theta}(\hat{x})+\frac{1}{\rho}\cos(\theta)\cos(\varphi)\frac{\partial f}{\partial\varphi}(\hat{x})\end{aligned}$$

$$\begin{aligned}\frac{\partial f}{\partial y}(\hat{x})&=\frac{\partial\left(f\circ g^{-1}\right)}{\partial y}(\hat{x})\\&=\operatorname{sen}(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial\rho}(\hat{x})+\frac{\cos(\theta)}{\rho\operatorname{sen}(\varphi)}\frac{\partial f}{\partial\theta}(\hat{x})+\frac{1}{\rho}\cos(\varphi)\operatorname{sen}(\theta)\frac{\partial f}{\partial\varphi}(\hat{x})\end{aligned}$$

y

$$\begin{aligned}\frac{\partial f}{\partial z}(\hat{x})&=\frac{\partial\left(f\circ g^{-1}\right)}{\partial z}(\hat{x})\\&=\cos(\varphi)\frac{\partial f}{\partial\rho}(\hat{x})-\frac{1}{\rho}\operatorname{sen}(\varphi)\frac{\partial f}{\partial\varphi}(\hat{x}).\end{aligned}$$

Como mencionamos en la sección anterior, con base en el teorema de la función inversa podemos dar la prueba del teorema de la función implícita, y lo más interesante es que también podemos hacer lo recíproco, es decir, probar el teorema de la función inversa a partir del teorema de la función implícita (lo que el lector hará en el problema 25 de este capítulo). Este hecho muestra que ambos teoremas son equivalentes, razón por la cual es suficiente dar la prueba de sólo uno de ellos.

Teorema 5.4 (de la función implícita) *Sean* $g_1, \ldots, g_m : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ *de clase* C^1 *en* U. *Si*

$$S = \left\{ (\hat{x}, \hat{y}) \in U \subset \mathbb{R}^m \times \mathbb{R}^k \mid g_1(\hat{x}, \hat{y}) = 0, \ldots, g_m(\hat{x}, \hat{y}) = 0 \right\}$$

y $(\hat{x}_0, \hat{y}_0) = \left(x_1^{(0)}, \ldots, x_m^{(0)}, y_1^{(0)}, \ldots, y_k^{(0)} \right) \in S$ *es tal que la matriz*

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

es invertible, entonces existen $\delta > 0$, $V \subset \mathbb{R}^k$ *un conjunto abierto, y una función* $h : V \subset \mathbb{R}^k \to \mathbb{R}^m$ *de clase* C^1 *en* V, *tales que* $\hat{y}_0 \in V$, $h(\hat{y}_0) = \hat{x}_0$ *y* $(h(\hat{y}), \hat{y}) \in S \cap B_\delta((\hat{x}_0, \hat{y}_0))$ *para toda* $\hat{y} \in V$. *Estas propiedades de* h *la hacen única.*

Demostración. Definamos $g : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^m$ como

$$g(\hat{x}, \hat{y}) = (g_1(\hat{x}, \hat{y}), \ldots, g_m(\hat{x}, \hat{y}))$$

y $f : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^m \times \mathbb{R}^k$ como

$$f(\hat{x}, \hat{y}) = (g(\hat{x}, \hat{y}), \hat{y}).$$

Es inmediato que f y g son de clase C^1 en U y que además:

1. como $(\hat{x}_0, \hat{y}_0) \in S$, entonces $f(\hat{x}_0, \hat{y}_0) = (g(\hat{x}_0, \hat{y}_0), \hat{y}_0) = (\hat{0}, \hat{y}_0)$, y

2. la matriz $Df(\hat{x}_0, \hat{y}_0)$, que está dada por

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) & \frac{\partial g_1}{\partial y_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial y_k}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) & \frac{\partial g_m}{\partial y_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial y_k}(\hat{x}_0, \hat{y}_0) \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}$$

es invertible.

Por tanto, por el teorema de la función inversa sabemos que existe $\delta > 0$ tal que:

1. f es inyectiva en

$$B_\delta(\hat{x}_0, \hat{y}_0) = \left\{(\hat{x}, \hat{y}) \in \mathbb{R}^m \times \mathbb{R}^k \mid \|(\hat{x}, \hat{y}) - (\hat{x}_0, \hat{y}_0)\| < \delta\right\} \subset U,$$

2. $f(B_\delta(\hat{x}_0, \hat{y}_0)) \subset \mathbb{R}^m \times \mathbb{R}^k$ es un conjunto abierto, y
3. $f^{-1} : f(B_\delta(\hat{x}_0, \hat{y}_0)) \subset \mathbb{R}^m \times \mathbb{R}^k$ es de clase C^1 en $f(B_\delta(\hat{x}_0, \hat{y}_0))$.

Definimos ahora $V \subset \mathbb{R}^k$ como

$$V = \left\{\hat{y} \in \mathbb{R}^k \mid (\hat{x}, \hat{y}) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0) \text{ para alguna } \hat{x} \in \mathbb{R}^m\right\}.$$

Nótese que $\hat{y}_0 \in V$, puesto que $(\hat{x}_0, \hat{y}_0) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0)$, y además se tiene que

$$f(S \cap B_\delta(\hat{x}_0, \hat{y}_0)) = \left\{\hat{0}\right\} \times V \subset \mathbb{R}^m \times \mathbb{R}^k. \tag{5.30}$$

En efecto, si $f(\hat{x}, \hat{y}) \in f(S \cap B_\delta(\hat{x}_0, \hat{y}_0))$, como $(\hat{x}, \hat{y}) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0)$ se tiene que $\hat{y} \in V$ y por lo tanto $f(\hat{x}, \hat{y}) = \left(\hat{0}, \hat{y}\right) \in \left\{\hat{0}\right\} \times V$. Por otra parte, si $\left(\hat{0}, \hat{y}\right) \in \left\{\hat{0}\right\} \times V$, como $\hat{y} \in V$, entonces existe $\hat{x} \in \mathbb{R}^m$ tal que $(\hat{x}, \hat{y}) \in S \cap B_\delta(\hat{x}_0, \hat{y}_0)$ de modo que $\left(\hat{0}, \hat{y}\right) = f(\hat{x}, \hat{y}) \in f(S \cap B_\delta(\hat{x}_0, \hat{y}_0))$.

De la identidad 5.30 obtenemos fácilmente que el conjunto $V \subset \mathbb{R}^m$ es abierto. En efecto, si $\hat{y} \in V$, entonces $\left(\hat{0}, \hat{y}\right) \in f(S \cap B_\delta(\hat{x}_0, \hat{y}_0)) \subset$

$f\left(B_\delta\left(\hat{x}_0,\hat{y}_0\right)\right)$, y como este último conjunto es abierto, existe $r>0$ tal que $B_r\left(\hat{0},\hat{y}\right)\subset f\left(B_\delta\left(\hat{x}_0,\hat{y}_0\right)\right)$. De esta forma, si $\hat{y}'\in B_r\left(\hat{y}\right)$ se tiene que

$$\left\|\left(\hat{0},\hat{y}'\right)-\left(\hat{0},\hat{y}\right)\right\|=\left\|\hat{y}'-\hat{y}\right\|<r,$$

de modo que $\left(\hat{0},\hat{y}'\right)\in B_r\left(\hat{0},\hat{y}\right)\subset f\left(B_\delta\left(\hat{x}_0,\hat{y}_0\right)\right)$, lo cual implica que existe $\hat{x}'\in\mathbb{R}^m$ tal que $\left(\hat{x}',\hat{y}'\right)\in B_\delta(\hat{x}_0,\hat{y}_0)$ y

$$\left(\hat{0},\hat{y}'\right)=f(\hat{x}',\hat{y}')=(g(\hat{x}',\hat{y}'),\hat{y}').$$

De lo anterior se deduce que $g(\hat{x}',\hat{y}')=\hat{0}$, de modo que $(\hat{x}',\hat{y}')\in S\cap B_\delta(\hat{x}_0,\hat{y}_0)$. De la definición de V se tiene que $\hat{y}'\in V$ y por lo tanto que $B_r\left(\hat{y}\right)\subset V$. Esta contención prueba que V es abierto.

De la identidad 5.30 y la inyectividad de f, también concluimos que para cada $\hat{y}\in V$ existe una única $\hat{x}\in\mathbb{R}^m$ tal que $\left(\hat{x},\hat{y}\right)\in S\cap B_\delta\left(\hat{x}_0,\hat{y}_0\right)$. En efecto, si $\hat{x},\hat{x}'\in\mathbb{R}^m$ son tales que $\left(\hat{x},\hat{y}\right),\left(\hat{x}',\hat{y}\right)\in S\cap B_\delta\left(\hat{x}_0,\hat{y}_0\right)$, entonces

$$f\left(\hat{x},\hat{y}\right),f\left(\hat{x}',\hat{y}\right)\in f\left(S\cap B_\delta\left(\hat{x}_0,\hat{y}_0\right)\right)=\left\{\hat{0}\right\}\times V$$

y por lo tanto $f\left(\hat{x},\hat{y}\right)=\left(\hat{0},\hat{y}\right)=f\left(\hat{x}',\hat{y}\right)$ de donde, por la inyectividad de f en la vecindad $B_\delta\left(\hat{x}_0,\hat{y}_0\right)$, se tiene que $\hat{x}=\hat{x}'$.

Con base en lo anterior, definimos $h:V\subset\mathbb{R}^k\to\mathbb{R}^m$ de la siguiente forma: dada $\hat{y}\in V$ hacemos $h\left(\hat{y}\right)=\hat{x}$, en donde $\hat{x}\in\mathbb{R}^m$ es el único elemento de $\mathbb{R}^m$ para el cual se satisface que $\left(\hat{x},\hat{y}\right)\in S\cap B_\delta\left(\hat{x}_0,\hat{y}_0\right)$. Como el lector podrá notar fácilmente, de la discusión anterior se tiene que h está bien definida y además es la única que se puede definir sobre el conjunto V con las propiedades de que $h\left(\hat{y}_0\right)=\hat{x}_0$, y $\left(h\left(\hat{y}\right),\hat{y}\right)\in S\cap B_\delta\left(\hat{x}_0,\hat{y}_0\right)$ para toda $\hat{y}\in V$.

Ahora sólo resta probar que h es de clase C^1 en V. Para ello, dado que la función f^{-1} sí está definida y es de clase C^1 sobre el conjunto $f\left(B_\delta\left(\hat{x}_0,\hat{y}_0\right)\right)$, si escribimos que $f^{-1}=\left(\left(f^{-1}\right)_1,\ldots,\left(f^{-1}\right)_{m+k}\right)$, bastará mostrar que

$$h\left(\hat{y}\right)=\left(\left(f^{-1}\right)_1\left(\hat{0},\hat{y}\right),\ldots,\left(f^{-1}\right)_m\left(\hat{0},\hat{y}\right)\right)$$

para cada $\hat{y}\in V$.

La identidad anterior se obtiene inmediatamente si observamos que para cada $\hat{y} \in V$ se cumple que

$$\begin{aligned}
&\left(h\left(\hat{y}\right), \hat{y}\right) \\
&= f^{-1}\left(f\left(h\left(\hat{y}\right), \hat{y}\right)\right) \\
&= f^{-1}\left(\left(\hat{0}, \hat{y}\right)\right) \\
&= \left(\left(f^{-1}\right)_1\left(\hat{0}, \hat{y}\right), \ldots, \left(f^{-1}\right)_m\left(\hat{0}, \hat{y}\right), \left(f^{-1}\right)_{m+1}\left(\hat{0}, \hat{y}\right), \ldots, \left(f^{-1}\right)_{m+k}\left(\hat{0}, \hat{y}\right)\right),
\end{aligned}$$

con lo cual terminamos nuestra prueba. ■

Con este teorema concluimos esta sección, este capítulo ¡y este texto!

5.6. Problemas

1. Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ derivable en cada punto $\hat{x} \in U$. Pruebe que:

   a) la gráfica de f (G_f) es una superficie suave en todos sus puntos

   b) si $\hat{x}_0 = (x_0, y_0, f(x_0, y_0)) \in G_f$, el plano tangente a G_f en $\hat{x}_0$ calculado de acuerdo con la definición 4.4 es el mismo que se obtiene si se calcula de acuerdo con la definición 5.2.

2. Calcule una ecuación cartesiana del plano tangente en un punto, de una superficie que está parametrizada por una función $\sigma : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ (definición 5.2).

3. Pruebe que la proposición 5.1 es independiente de las funciones coordenadas de f que se tomen.

4. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ derivable en el punto $\hat{x}_0 \in U$. Pruebe que existen $r > 0$ y $M > 0$ tales que

   $$\frac{\|f(\hat{x}) - f(\hat{x}_0)\|}{\|\hat{x} - \hat{x}_0\|} \leq M$$

   para toda $\hat{x} \in B_r(\hat{x}_0) \setminus \{\hat{x}_0\} \subset U$.

5. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $g : V \subset \mathbb{R}^m \to \mathbb{R}^k$ y $\hat{x}_0 \in U$ tales que $f(U) \subset V$. Definimos $\varphi : U \subset \mathbb{R}^n \to \mathbb{R}^k$ como

$$\varphi(\hat{x}) = \begin{cases} \frac{g(f(\hat{x}))-g(f(\hat{x}_0))-Dg(f(\hat{x}_0))(f(\hat{x})-f(\hat{x}_0))}{\|f(\hat{x})-f(\hat{x}_0)\|} & \text{si } f(\hat{x}) - f(\hat{x}_0) \neq \hat{0} \\ \hat{0} & \text{si } f(\hat{x}) - f(\hat{x}_0) = \hat{0} \end{cases}$$

Pruebe que, si f es continua en $\hat{x}_0$ y g es derivable en $\hat{y}_0 = f(\hat{x}_0)$, entonces φ es continua en $\hat{x}_0$.

6. Sean $f : U \subset \mathbb{R}^n \to \mathbb{R}$ y $g : U \subset \mathbb{R}^n \to \mathbb{R}^m$ derivables en el punto $\hat{x}_0 \in U$. Pruebe que la función $(fg)(\hat{x}) = f(\hat{x})g(\hat{x})$ es derivable en $\hat{x}_0 \in U$ y dé una fórmula para la $D(fg)(\hat{x}_0)$.

7. Si $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \in M_{1\times 3}(\mathbb{R})$ y

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ b_{21} & \cdots & b_{2n} \\ b_{31} & \cdots & b_{3n} \end{bmatrix} \in M_{3\times n}(\mathbb{R}),$$

definimos el producto cruz de la matriz A por la matriz B (que denotaremos por $A \times B$), como la matriz de $3 \times n$ (con entradas reales) cuyas entradas de su $j-$ésima columna coinciden con las coordenadas del vector

$$\begin{aligned}(a_{11}, a_{12}, a_{13}) &\times (b_{1j}, b_{2j}, b_{3j}) \\ &= (a_{12}b_{3j} - a_{13}b_{2j}, a_{13}b_{1j} - a_{11}b_{3j}, a_{11}b_{2j} - a_{12}b_{1j})\end{aligned}$$

es decir,

$$A \times B := \begin{bmatrix} a_{12}b_{31} - a_{13}b_{21} & \cdots & a_{12}b_{3n} - a_{13}b_{2n} \\ a_{13}b_{11} - a_{11}b_{31} & \cdots & a_{13}b_{1n} - a_{11}b_{3n} \\ a_{11}b_{21} - a_{12}b_{11} & \cdots & a_{11}b_{2n} - a_{12}b_{1n} \end{bmatrix} \in M_{3\times n}(\mathbb{R}).$$

Sean $f = (f_1, f_2, f_3)$, $g = (g_1, g_2, g_3) : U \subset \mathbb{R}^n \to \mathbb{R}^3$ derivables en el punto $\hat{x}_0 \in U$. Pruebe que la función $(f \times g)(\hat{x}) := f(\hat{x}) \times g(\hat{x})$ es derivable en $\hat{x}_0 \in U$ y dé una fórmula para $D(f \times g)(\hat{x}_0)$ en términos del producto cruz de matrices definido en el párrafo anterior.

8. Encuentre todas las funciones $f : \mathbb{R}^n \to \mathbb{R}^n$ tales que $Df(\hat{x})$ es una matriz diagonal para toda $\hat{x} \in \mathbb{R}^n$. Pruebe su respuesta.

9. Abusar de la notación (como de cualquier otra cosa) suele causar problemas. En particular, usar letras (que casi siempre denotan variables) para referirse también a funciones nos puede llevar a errores. Sea $w = f(x, y, z)$ y $z = g(x, y)$. Por la regla de la cadena, se tiene que:

$$\frac{\partial w}{\partial x} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial x} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial x}.$$

Como x y y son variables independientes, entonces $\frac{\partial y}{\partial x} = 0$ y como $\frac{\partial x}{\partial x} = 1$, se tiene que:

$$\frac{\partial w}{\partial x} = \frac{\partial w}{\partial x} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial x}$$

Así, $\frac{\partial w}{\partial z}\frac{\partial z}{\partial x} = 0$. Si $f(x, y, z) = 2x + y + 3z$ y $g(x, y) = 5x + 18$, entonces $\frac{\partial w}{\partial z} = 3$ y $\frac{\partial z}{\partial x} = 5$ y por lo tanto $0 = 15$. ¿Cuál es el error?

10. Sean f y g definidas como

$$f(u, v) = (u\cos(v), u\,\mathrm{sen}(v))$$

con $0 < u$ y $-\pi/2 < v < \pi/2$, y

$$g(x, y) = (\sqrt{x^2 + y^2}, \arctan(y/x))$$

con $0 < x$. Calcule $D(f \circ g)(x, y)$ y $D(g \circ f)(u, v)$.

11. *a*) Suponga que la variable w está en función de las variables x, y, z y t (es decir: $w = f(x, y, z, t)$), que $x = g(u, z, t)$ y que $z = h(u, t)$. Tomando en cuenta todas estas relaciones, calcule $\frac{\partial w}{\partial t}$.

b) Si

$$\begin{aligned} f(x, y, z, t) &= 2xy + 3z + t^2 \\ g(u, z, t) &= ut\,\mathrm{sen}(z) \\ h(u, t) &= 2u + t, \end{aligned}$$

calcule $\frac{\partial w}{\partial t}$ para $u = 1, t = 2$ y $y = 3$.

12. Suponga que $f : \mathbb{R}^2 \to \mathbb{R}$ es tal que

$$\frac{\partial f}{\partial x}(2,1) = 3, \qquad \frac{\partial f}{\partial y}(2,1) = -2,$$

$$\frac{\partial^2 f}{\partial x^2}(2,1) = 0, \qquad \frac{\partial^2 f}{\partial y^2}(2,1) = 2,$$

$$\frac{\partial^2 f}{\partial y \partial x}(2,1) = 1, \qquad \frac{\partial^2 f}{\partial x \partial y}(2,1) = 1.$$

Si $g(u,v) = (u^2 + v, uv)$, calcule $\frac{\partial^2 (f \circ g)}{\partial v \partial u}(1,1)$.

13. Sean $f, g : U \subset \mathbb{R}^2 \to \mathbb{R}$ de clase C^1 en U. Definimos $F : U \subset \mathbb{R}^2 \to \mathbb{R}^2$ y $H : U \subset \mathbb{R}^2 \to \mathbb{R}$ como

$$F(x,y) = (f(x,y), g(x,y)) \qquad \text{y} \qquad H(x,y) = \|F(x,y)\|^2 .$$

Demuestre que no existe $\hat{x}_0 \in U$ que satisfaga las siguientes dos propiedades:

a) H tiene un máximo local en $\hat{x}_0$

b) $DF(\hat{x}_0)$ es invertible.

14. Sea $f : \mathbb{R}^n \to \mathbb{R}^n$ derivable en un punto $\hat{x}_0 \in \mathbb{R}^n$ y además supóngase que $\hat{x}_0$ es un punto fijo de f (es decir: $f(\hat{x}_0) = \hat{x}_0$). Si A denota a la matriz $Df(\hat{x}_0)$ y $k \in \mathbb{N}$, encuentre una función $g : \mathbb{R}^n \to \mathbb{R}^n$ derivable en $\hat{x}_0$ tal que $g(\hat{x}_0) = \hat{x}_0$ y $Dg(\hat{x}_0) = A^k$.

15. Sea $f : U \subset \mathbb{R}^3 \to \mathbb{R}$ una función que está expresada en términos de las coordenadas cartesianas (x,y,z) de cada punto $\hat{x} \in U$. Use la función de cambio de coordenadas cilíndricas (ρ, θ, z) a coordenadas cartesianas (x,y,z), y la regla de la cadena, para encontrar (en cada punto $\hat{x} \in U$) una base ortonormal de $\mathbb{R}^3$ en la cual se pueda expresar a la derivada de f en $\hat{x}$ $(Df(\hat{x}))$ en términos de las derivadas parciales $\frac{\partial f}{\partial \rho}$, $\frac{\partial f}{\partial \theta}$ y $\frac{\partial f}{\partial z}$. Compare con lo obtenido en el problema 23 del capítulo 4.

16. Repita el problema anterior ahora para las coordenadas esféricas (ρ, θ, φ) y compare con lo obtenido en el problema 24 del capítulo 4.

17. Sea $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ de clase C^2 en U tal que $\frac{\partial f}{\partial x}(x, y) > 0$ para toda $(x, y) \in U$, y sea $(x_0, y_0) \in U$. Si $I = \{y \in \mathbb{R} \,|(x_0, y) \in U\}$, definimos $g : I \subset \mathbb{R} \to \mathbb{R}$ como $g(y) = f(x_0, y)$. Pruebe que:

    a) existe $\delta > 0$ y $h : (y_0 - \delta, y_0 + \delta) \subset I \to \mathbb{R}$ de clase C^1 tal que $f(h(y), y) = f(x_0, y_0)$ para toda $y \in (y_0 - \delta, y_0 + \delta)$

    b) g tiene un máximo (mínimo) local en y_0 si y sólo si h tiene un mínimo (máximo) local en y_0. (suponga que $\frac{\partial^2 f}{\partial y^2}(x_0, y_0) \neq 0$, para que todo sea más fácil). Interprete geométricamente.

18. Sean $g_1, \ldots, g_m : U \subset \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$, S y $(\hat{x}_0, \hat{y}_0) \in S$ como en el teorema de la función implícita. Si

$$A = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

y

$$B = \begin{bmatrix} \frac{\partial g_1}{\partial y_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_1}{\partial y_k}(\hat{x}_0, \hat{y}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial y_1}(\hat{x}_0, \hat{y}_0) & \cdots & \frac{\partial g_m}{\partial y_k}(\hat{x}_0, \hat{y}_0) \end{bmatrix}$$

pruebe que:

$$Dh(\hat{y}_0) = -A^{-1}B,$$

donde h es la función cuya existencia es garantizada en el mencionado teorema.

19. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^1 en U y $S = \{\hat{x} \in U \mid f(\hat{x}) = cte\}$. Si $\hat{x} \in \mathbb{R}^n$, denotamos por $\hat{x}^{(-i)}$ el elemento de $\mathbb{R}^{n-1}$ que se obtiene de $\hat{x}$ al "eliminarle" su $i-$ésima coordenada, con $i \in \{1, \ldots, n\}$. Sea $\hat{x}_0 = (x_0^{(1)}, \ldots, x_0^{(n)}) \in S$.

a) Pruebe que, si $\frac{\partial f}{\partial x_i}(\hat{x}_0) \neq 0$, entonces existe $h_i : U_i \subset \mathbb{R}^{n-1} \to \mathbb{R}$ de clase C^1 en U_i tal que $\hat{x}_0^{(-i)} \in U_i$, $h_i(\hat{x}_0^{(-i)}) = x_0^{(i)}$ y

$$(x_1, \ldots, x_{i-1}, h_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n), x_{i+1}, \ldots, x_n) \in S$$

para todo $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in U_i$.

b) Si h_i es la función del inciso anterior, calcule $\frac{\partial h_i}{\partial x_{i+1}}(\hat{x}_0^{(-i)})$ (para $i = n$ calcule $\frac{\partial h_n}{\partial x_1}(\hat{x}_0^{(-n)})$) en términos de derivadas parciales de f.

c) Si $\frac{\partial f}{\partial x_i}(\hat{x}_0) \neq 0$ para cada $i \in \{1, \ldots, n\}$, pruebe que

$$\frac{\partial h_1}{\partial x_2}(\hat{x}_0^{(-1)})\frac{\partial h_2}{\partial x_3}(\hat{x}_0^{(-2)}) \cdots \frac{\partial h_{n-1}}{\partial x_n}(\hat{x}_0^{(-(n-1))})\frac{\partial h_n}{\partial x_1}(\hat{x}_0^{(-n)}) = (-1)^n.$$

20. Sea $g : U \subset \mathbb{R}^n \to \mathbb{R}$ de clase C^1 (en U) y $\hat{x}_0 \in U$ tal que $g(\hat{x}_0) = 0$ y $\frac{\partial g}{\partial x_n}(\hat{x}_0) \neq 0$. Si $\hat{v} \neq \hat{0}$ es tal que $\hat{v} \cdot \nabla g(\hat{x}_0) = 0$, pruebe que existen $\delta > 0$ y

$$\gamma : (-\delta, \delta) \subset \mathbb{R} \to \mathbb{R}^n$$

derivable tales que $g(\gamma(t)) = 0$ para toda $t \in (-\delta, \delta)$, $\gamma(0) = \hat{x}_0$ y $\gamma'(0) = \hat{v}$. Interprete geométricamente.

21. Sea $g : U \subset \mathbb{R}^3 \to \mathbb{R}$ de clase C^1 (en U) y $\hat{x}_0 = (x_0, y_0, z_0) \in U$ tal que $g(\hat{x}_0) = 0$ y $\frac{\partial g}{\partial z}(\hat{x}_0) \neq 0$. Pruebe que:

a) existe $V \subset \mathbb{R}^2$ abierto, y $\sigma : V \subset \mathbb{R}^2 \to \mathbb{R}^3$ de clase C^1 en V, tales que $(x_0, y_0) \in V$ y $g(\sigma(x, y)) = 0$ para toda $(x, y) \in V$. Interprete geométricamente,

b) el plano tangente al conjunto de nivel 0 de g ($N_0(g)$) calculado usando la parametrización σ del inciso anterior es el mismo si se calcula usando la definición 4.7.

22. Sean $g_1, g_2 : U \subset \mathbb{R}^3 \to \mathbb{R}$ de clase C^1 (en U) y $\hat{x}_0 = (x_0, y_0, z_0) \in U$ tales que $g_1(\hat{x}_0) = g_2(\hat{x}_0) = 0$ y $\nabla g_1(\hat{x}_0) \times \nabla g_2(\hat{x}_0) \neq \hat{0}$. Pruebe que:

a) existe $\delta > 0$ y
$$\gamma : (-\delta, \delta) \subset \mathbb{R} \to \mathbb{R}^3$$
derivable tal que $g_1(\gamma(t)) = g_2(\gamma(t)) = 0$ para toda $t \in (-\delta, \delta)$, $\gamma(0) = \hat{x}_0$ y $\gamma'(0) \neq \hat{0}$,

b) la recta tangente en el punto $\gamma(0)$ a la curva descrita por γ, es la misma recta que se obtiene al intersectar al plano tangente en el punto $\hat{x}_0$ del conjunto de nivel 0 de la función g_1, con el plano tangente en el punto $\hat{x}_0$ del conjunto de nivel 0 de la función g_2. Interprete geométricamente.

23. Considere el conjunto de soluciones de las ecuaciones:
$$\begin{aligned} 2x + y + 2z + u - v - 1 &= 0 \\ xy + z - u + 2v - 1 &= 0 \\ yz + xz + u^2 + v &= 0 \end{aligned}$$

a) Muestre que, en una vecindad del punto $(1, 1, -1, 1, 1)$, las variables x, y y z (del conjunto de soluciones) se pueden poner en función de las variables u y v.

b) Calcule la derivada de la función del inciso anterior en el punto $(1, 1)$.

c) Encuentre todas la ternas de variables que se puedan poner en función de las restantes dos, en una vecindad del mismo punto.

24. Sea
$$f(x) = \begin{cases} x^2 \operatorname{sen}(1/x) + \frac{x}{2} & \text{si } x \neq 0 \\ \\ 0 & \text{si } x = 0 \end{cases}$$

a) calcule $f'(x)$ para toda $x \in \mathbb{R}$

b) pruebe que para toda $\delta > 0$ existen $x, y \in (-\delta, \delta)$ tales que $f'(x) < 0$ y $f'(y) > 0$

c) pruebe que f no es invertible en ninguna vecindad del cero. ¿Este ejemplo contradice el Teorema de la Función Inversa?

25. Pruebe el teorema de la función inversa a partir del teorema de la función implícita.

26. Sean $f : \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 y $c > 0$ tales que para toda $\hat{x}, \hat{y} \in \mathbb{R}^n$ se tiene que
$$\|f(\hat{x}) - f(\hat{y})\| \geq c\,\|\hat{x} - \hat{y}\|\,.$$
Pruebe que:

    a) para toda $\hat{x} \in \mathbb{R}^n$, $Df(\hat{x})$ es invertible
    b) $f(\mathbb{R}^n)$ es abierto
    c) $f(\mathbb{R}^n)$ es cerrado
    d) f es biyectiva y $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ es de clase C^1 (en $\mathbb{R}^n$).

27. Sea $g = (g_1, \ldots, g_m) : A \subset \mathbb{R}^n \to \mathbb{R}^m$ de clase C^1 (en A) con $n > m$. Sea $\hat{x}_0 \in A$ tal que $g(\hat{x}_0) = \hat{0}$.

    a) Si la matriz
    $$M = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial g_1}{\partial x_m}(\hat{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\hat{x}_0) & \cdots & \frac{\partial g_m}{\partial x_m}(\hat{x}_0) \end{bmatrix}$$
    es de rango m, pruebe que existen $U, V \subset \mathbb{R}^n$ abiertos y $f : U \to V$ una biyección de clase C^1 en U (y f^{-1} de clase C^1 en V) tales que $\hat{x}_0 \in V \subset A$ y $g(f(x_1, \ldots, x_n)) = (x_1, \ldots, x_m)$ para todo $(x_1, \ldots, x_n) \in U$.
    b) Describa el conjunto $(g \circ f)^{-1}(\hat{0})$.
    c) Use el primer inciso para demostrar que si $Dg(\hat{x}_0)$ tiene rango máximo (m), entonces $Dg(\hat{x})$ tiene rango máximo para $\hat{x}$ en una vecindad de $\hat{x}_0$.
    d) Si $Dg(\hat{x}_0)$ no tiene rango máximo (es decir, si el rango de $Dg(\hat{x}_0)$ es $k < m$) ¿se sigue cumpliendo un resultado análogo al del primer inciso? ¿al del cuarto inciso? (es decir, si $Dg(\hat{x}_0)$ tiene rango $k < m$ ¿existe una vecindad de $\hat{x}_0$ en la que Dg sigue teniendo rango k?). Pruebe sus respuestas.

e) Interprete geométricamente.

28. Sea $g : U \subset \mathbb{R}^n \to \mathbb{R}^m$ de clase C^1 (en U) con $n > m$. Pruebe que g no es inyectiva.

29. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ de clase C^1 (en U) con $n < m$.

    a) Si $\hat{x}_0 = \left(x_1^{(0)}, \ldots, x_n^{(0)}\right) \in U$ es tal que $Df(\hat{x}_0)$ tiene rango n (como matriz de $m \times n$), pruebe que existen $\tilde{U}, \tilde{V} \subset \mathbb{R}^m$ abiertos y $h : \tilde{U} \to \tilde{V}$ una biyección de clase C^1 en $\tilde{U}$ (y h^{-1} de clase C^1 en $\tilde{V}$) tales que $\left(x_1^{(0)}, \ldots, x_n^{(0)}, 0, \ldots, 0\right) \in \tilde{V}, f(\hat{x}_0) \in \tilde{U}$ y $h(f(x_1, \ldots, x_n)) = (x_1, \ldots, x_n, 0, \ldots, 0) \in \mathbb{R}^m$ para todo $(x_1, \ldots, x_n) \in f^{-1}(\tilde{U})$ tal que $(x_1, \ldots, x_n, 0, \ldots, 0) \in \tilde{V}$.

    b) Use el primer inciso para demostrar que si $Df(\hat{x}_0)$ tiene rango máximo (n), entonces $Df(\hat{x})$ tiene rango máximo para $\hat{x}$ en una vecindad de $\hat{x}_0$.

    c) Si $Df(\hat{x}_0)$ no tiene rango máximo (es decir, si el rango de $Df(\hat{x}_0)$ es $k < n$) ¿se sigue cumpliendo un resultado equivalente al del primer inciso? ¿al del tercer inciso? (es decir, si $Df(\hat{x}_0)$ tiene rango $k < n$ ¿existe una vecindad de $\hat{x}_0$ en la que Df sigue teniendo rango k?). Pruebe sus respuestas.

    d) Interprete geométricamente.

30. Sea $f(x, y) = (e^x \cos(y), e^x \operatorname{sen}(y))$.

    a) Pruebe que f no es inyectiva en $\mathbb{R}^2$.

    b) Pruebe que para cualquier $\hat{x} = (x, y) \in \mathbb{R}^2$ existe $\delta > 0$ tal que f es invertible en $B_\delta(\hat{x})$.

    c) ¿Existe una función de $\mathbb{R}$ en $\mathbb{R}$ que tenga las dos propiedades anteriores? Pruebe su respuesta.

31. Sean $f, g : U \subset \mathbb{R}^3 \to \mathbb{R}$ de clase C^1 en U. Pruebe que la función

$$F(x, y, z) = (f(x, y, z), g(x, y, z), f(x, y, z) + g(x, y, z))$$

es tal que, si tuviera inversa en alguna vecindad de algún punto de U, ésta no sería derivable.

32. Sea $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ de clase C^1 (en U) y $A \subset U$. Pruebe que, si $\det(Df(\hat{x})) \neq 0$ para toda $\hat{x} \in int(A)$, entonces $f(int(A)) \subset int(f(A))$.

33. Sea $f : U \subset \mathbb{R}^3 \to \mathbb{R}$ de clase C^2 en U. Si f está expresada en términos de coordenadas esféricas, calcule $\frac{\partial^2 f}{\partial z^2}(\hat{x})$ en términos de estas mismas coordenadas (*sugerencia*: proceda como en el ejemplo 5.4).

Bibliografía

[1] Spivak, M., *Calculus*, 3a. edición. Editorial Reverté, Barcelona, 2012. 682 pp.

[2] Friedberg, Stephen H., Insen, Arnold, J., Spence, Lawrence, E. *Linear algebra*, 4th ed. PHI Learning, New Delhi, 2013. 601 pp.

[3] Swokowski, E. W., *Cálculo con geometría analítica*, 2a. edición. Grupo Editorial Iberoamerica, México, 1998. 1097 pp.

Índice analítico

acumulación
    punto de, 43
afín
    función, 268
aislado
    punto, 43
arco
    longitud de, 202
    parametrización por longitud de, 203

base
    canónica, 238
    ortonormal, 239
bola
    con centro en un punto en $\mathbb{R}^n$, 31
Bolzano-Weierstrass
    teorema de, 49

cadena
    regla de la, 194, 257, 303, 307, 430
Cauchy
    criterio de convergencia de, 116
    sucesión de, 116
Cauchy-Schwarz
    desigualdad de, 22
    desigualdad, de, 393
cerradura
    de un conjunto, 41
cicloide, 182
cilíndricas
    coordenadas, 76
circunferencia
    osculadora, 210
coeficiente multinomial, 406
compacto
    conjunto, 154
conexo
    conjunto, 58
conjunto
    abierto, 30, 37
    acotado, 49
    cerrado, 37
    cerradura de un, 41
    compacto, 154
    conexo, 58
    convexo, 61

de nivel, 100
disconexo, 57
estrellado, 94
exterior, 34
frontera, 34
imagen directa de un, 104
imagen inversa de un, 103
interior, 34
parametrización de un, 177
conjuntos
separados, 57
convexo
conjunto, 61
coordenada
función, 96
sucesión, 113
coordenadas, 9
cilíndricas, 76
esféricas, 80
polares, 70
cuadrados mínimos
método de los, 408
cubierta
abierta, 154
curva, 105
curvatura de una, 210
definición de, 198
longitud de una, 202
regular, 198
reparametrización de una, 204
suave, 198
torsión de una curva, 214, 215
curvatura, 210
centro de, 211
radio de, 211

derivada
de una función de $\mathbb{R}$ en $\mathbb{R}^n$, 184
de una función de $\mathbb{R}^n$ en $\mathbb{R}$, 269
de una función de $\mathbb{R}^n$ en $\mathbb{R}^m$, 412
direccional, 250
en coordenadas polares, 324
parcial, 263
cruzada, 331
desigualdad
de Cauchy-Schwarz, 22, 393
de Hölder, 410
de Minkowski, 410
direccional
derivada, 250
directa
imagen, 104
disconexo
conjunto, 57
distancia, 14
euclideana, 29

esféricas
coordenadas, 80
euclideana
distancia, 29
norma, 13
evoluta, 213
exterior
de un conjunto, 34
punto, 34

forma cuadrática, 358

no degenerada, 359
seminegativa, 359
semipositiva, 359
formas cuadráticas, 374
Frenet-Serret
fórmulas de, 218
frontera
de un conjunto, 34
punto, 34
función
lineal, 169
función
afín, 268
coordenada, 96
de clase C^k, 337
gráfica de una, 98
mínimo de una, 352
mínimo local de una, 352
máximo de una, 352
máximo local de una, 352
matriz hessiana de una, 359

gráfica
de una función, 98
gradiente
en coordenadas polares, 325
vector, 281, 283

Hölder
desigualdad de, 410
hessiana
de una función, 359

imagen
directa, 104
inversa, 103
implícita
teorema de la función, 439, 457
interior
de un conjunto, 34
punto, 34
inversa
teorema de la función, 473

Lagrange
multiplicadores de, 388
teorema de los multiplicadores de, 388
lineal
función, 169
longitud
de arco, 202
parametrización por, 203
de una curva, 202

mínimo
de una función, 352
local
de una función, 352
máximo
de una función, 352
local
de una función, 352
matriz
hessiana, 358
jacobiana, 424
ortonormal, 244
Minkowski
desigualdad de, 410

multiplicadores de Lagrange, 388
    prueba del teorema de los, 459

norma
    euclideana, 13
    infinito, 26
    uno, 26

ortonormal
    base, 239
    matriz, 244

parametrización
    de un conjunto, 177
    de una superficie, 417
    por longitud de arco, 203, 205
parcial
    derivada, 263
plano
    osculador, 210
    tangente, 276, 310, 414, 417
polares
    coordenadas, 70
poligonal, 63
polinomio
    de Taylor, 345
producto
    interior, 20
    punto, 20
punto
    aislado, 43
    crítico, 354
    de acumulación, 43
    exterior, 34
    frontera, 34
    interior, 34
    silla, 356

rectángulos anidados
    teorema de los, 51
recta
    tangente, 310
recta tangente
    a una curva, 185
regla
    de la cadena, 194, 257, 303, 307, 430
reparametrización, 204

sucesión
    de imágenes, 123
sucesión
    de Cauchy, 116
sucesión
    acotada, 120
    coordenada, 113
    en $\mathbb{R}^n$, 112
    rango de una, 120
superficie, 105, 417
    parametrización de una, 417
    suave, 417
superficies, 414

tangente
    plano, 276, 310, 414, 417
    recta, 310
Taylor
    polinomio de, 345
    Teorema de, 345
teorema

de Bolzano-Weierstrass, 49
de la función implícita, 439, 457
prueba del, 485
de la función inversa, 473
prueba del, 474
de los multiplicadores de Lagrange, 388
prueba del teorema, 459
de los rectángulos anidados, 51
de Taylor, 345
torsión
de una curva suave, 214, 215

vecindad
agujerada, 43
de un punto en $\mathbb{R}^n$, 30
vector
binormal unitario, 211
gradiente, 281, 283
normal unitario, 210
tangente unitario, 210

Cálculo diferencial de varias variables
editado por la Universidad Nacional Autónoma de México
se terminó de imprimir el 15 de enero de 2019
en los talleres de Amy Soluciones Gráficas S. A. de C. V.
Corregidora 79. Santa Anita Iztacalco.
CP. 8300. Ciudad de México

El tiraje fue de 1000 ejemplares

Impresión offset sobre papel Book creamy de 60 g.
En su composición se utilizó tipografía
Computer modern de 11/13 pts.

Impresión offset

El cuidado de la edición estuvo a cargo de
Patricia Magaña Rueda